1. INFO 202 - Course Overview

INFO 202 - 27 August 2008

Bob Glushko

Plan for INFO 202 Lecture #1

Course Overview
Course Topics and Issues
Course Administrivia
The One Minute Course

It is important to think ABSTRACTLY about information and the conceptual tasks of organizing and finding it.

We experience information as individuals, in association with other individuals, or as part of a business or business ecosystem.

We must recognize the profound impact of new technologies and their co-evolution with IO and IR techniques.

But we must still appreciate the "classical" knowledge that is often ignored by those chasing new technologies.
Why This Course Exists

Because the IO and IR fields are changing rapidly and are critical in much of what we do
To build a foundation and framework for much of your I-School experience
To establish a baseline of knowledge for students with heterogeneous backgrounds

Why This Course Is Challenging / Unique / 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 / content IN a discipline to studying information / content AS a discipline

You must learn to look past the presentation / rendition / technology reification / thinginess of information to see it more abstractly as structure and meaning

The diversity of perspectives and backgrounds in your class will be one of the challenges in this course
### IO Buzzwords

<table>
<thead>
<tr>
<th>Semantics</th>
<th>Syntax</th>
<th>Semiotics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorization</td>
<td>Classification</td>
<td>&quot;the Work&quot;</td>
</tr>
<tr>
<td>Taxonomy</td>
<td>Folksonomy</td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>Tag</td>
<td>Web 2.0</td>
</tr>
<tr>
<td>Metadata</td>
<td>Metamodel</td>
<td>Facet</td>
</tr>
<tr>
<td>Conceptual Model</td>
<td>Schema</td>
<td>Microformat</td>
</tr>
<tr>
<td>Ontology</td>
<td>Information Architecture</td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td>Mash-up</td>
<td></td>
</tr>
<tr>
<td>Interoperability</td>
<td>Compliance</td>
<td>Standardization</td>
</tr>
</tbody>
</table>

### IR Buzzwords

<table>
<thead>
<tr>
<th>Collection</th>
<th>Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Browsing</td>
<td>“Berrypicking”</td>
</tr>
<tr>
<td>Recall</td>
<td>Precision</td>
</tr>
<tr>
<td>Link</td>
<td>Anchor</td>
</tr>
<tr>
<td>Crawling</td>
<td>Data Mining</td>
</tr>
<tr>
<td>Boolean</td>
<td>Bayesian</td>
</tr>
<tr>
<td>Term Weighting</td>
<td>TF/IDF</td>
</tr>
<tr>
<td>Vectors</td>
<td>Dimensionality Reduction</td>
</tr>
</tbody>
</table>
What Is Information?

The content of libraries, most of which is in text format and whose "catalog" descriptions are in the same format

Descriptions of any thing or object that can have information associated with it

Non-text or multimedia information, whose description is not in the same format

Information found in databases or business applications

DOCUMENT = "a purposeful and self-contained collection of information"

Why Do People Search for Information?

People search to find information that lets them make better decisions or act more effectively

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"
How Do People Search for Information?

How people search depends on what they're searching for and where they're searching for it.

Sometimes people search for specific answers.

Sometimes they don't know what they're looking for.

They refine their questions and strategy on the basis of what they find.

Their actions are strongly shaped by the way the information resources they're searching are structured, and how what they find is presented to them.

How People "Find Out About"
How "Finding Out About" is Realized in IR Systems

How Can We Organize Information?

What does it mean to organize information?

How is organizing information like and unlike organizing things?

We organize when we assign words to things or concepts

We also organize when we use categories and classifications... but every classification system is biased
Rabbit or Duck?

What is Meaning? Where is Meaning?

What does it mean to name something?

How do others understand what we mean?
Defining What Something Means

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

Sorry Pluto
Why Is Organizing Information Difficult?

People use different words for the same things, and the same words for different things.

Organization is always (explicitly or implicitly) being done in some context.

Every context produces different biases in naming and categorization.
What is the Context / Scope of Information Organization?

- Individual concepts and information components
- Documents and databases
- Personal collection
- Within a group / community / enterprise
- Between enterprises
- "In the world"

Who Organizes Information?

- Professionals (the emphasis of traditional library science)
- Authors
- Users ("folksonomy" and other end-user "tagging")
- Machines (computer programs)
What Principles or Best Practices Govern Information Organization?

Traditional library science  
Data model (aka database) design; systems analysis  
Document type design  
Information architecture  
Cognitive science  
Business Operations and Strategy  
Economics

What are the Roles of IO and IR in Business?

Information technology has radically transformed the nature of business models and business architecture  
We are moving from information IN business to information AS business - to an information and services economy  
Some businesses are inherently about information organization, management and distribution  
Every large business has important new Information management concerns
A Multidimensional Categorization Problem

A Tour of the Syllabus

Foundations followed by IO followed by IR

Progressively more rigorous view of what meaning is

Progressively larger scope in which we are trying to organize information and specify meaning

Theory interleaved with applications
Textbooks and Readings


IR book: Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval* (2008, hot off the presses)

Readings come from library science, computer science, cognitive science, linguistics, philosophy, systems analysis, information architecture, business case studies

Course Reader(s) available at Copy Central, 2560 Bancroft

Assignments and Exams

Eight assignments

Midterm 27 October (regular class meeting time)

Final 15 December (9am-Noon)
Weekly Discussion Section

Candidate section times (we will select 3):

- M 11-12
- M 4-5
- M 5-6
- M 6-7
- Tu 11-12

Grading

Assignments 40%
Midterm exam 20%
Final exam 30%
In class/section and online participation 10%
List Serve

Sign up for "i202" class list serve

- e-mail to majordomo@ischool.berkeley.edu
- Subject: Leave blank
- Body of message: subscribe i202

Course Blog

http://blogs.ischool.berkeley.edu/i202f08/

You'll be submitting your first assignment as a post to this blog

Gives you a chance to demonstrate "insight" outside of classroom
Class Policies and Expectations

My office hours are M 4-5 and W 11-12 or by appt

My lecture notes will be posted by 8:45am the day of the lecture

You should do the assigned reading before each class so that you can participate

You and I will monitor and manage your signal/noise and that of others

Your cell phone can ring in class no more than once during the semester

---

Today's Assigned Readings

Vannevar Bush, *As We May Think*

Gemmell, Bell, & Lueder, *My Life Bits*

Jorge Luis Borges, *The Library of Babel*

Ramana Rao, *From IR to Search and Beyond*

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, *Introduction to Information Retrieval*, Section 19.1
Readings for Wednesday 3 September

Malcolm Wheatley, "Operation Clean Data," CIO (July 2004)


R. Baron, et al., "Electronic Health Records: Just around the Corner? Or over the Cliff?," Annals of Internal Medicine, (August 2005)


Assignment 1

Find a "202 in the News" story (published since 1 January 2007)

Post to the blog a bibliographic citation or URL and a 100-word summary

Assign at least 4 "tags" to the story from the "IO and IR Buzzwords" in today's lecture notes

Identify the lecture in the syllabus where the story seems to fit
Class of 2010 202 Party

Tuesday, September 9, 6-9pm
At my house in the Berkeley hills, 34 Stephens Way
No public transportation, so this will encourage you to talk to each other