

I 202: INFORMATION ORGANIZATION & RETRIEVAL FALL 2025

Class 1: Introduction

Today's Outline

Instructor Introductions

Student Introductions

Information vs Data

Course Structure



INSTRUCTOR: MARTI HEARST

I School and CS Professor

Research: Information Visualization, Natural Language Processing, HCI

Fav Berkeley Restaurant: Zachary's Pizza, Great China, Millennium



TA: SUNNY LEE

MIMS '26

Bachelors in Neuroscience & Cognitive Science

Work experience in healthcare & academic research

Current interests: UX research, data science, product mgmt



READER: SARAH ALGASHGARI

MIMS '26

IMSA President

Information Visualization, Consulting

STUDENT INTRODUCTIONS

Introduce yourself to your neighbor

STUDENT BACKGROUND: RESULTS OF SURVEY

Where Did We Arrive From?



Favorite Berkeley Restaurants

Mezzo

Saigon Express

Imm Thai

Saul's Deli

Noodle Dynasty

The Smokehouse

Jot Mahal

HanKki

Kura Sushi

Lulu

Masa Raman

What Excites You about the I School

Meeting other data science and machine learning enthusiast.

Interdisciplinary nature

The courses, the community, and the interdisciplinary aspect of the I School

The interdisciplinary nature of it

It's diversity and expert professors in the field

Cool research + great vibes and people

The diversity of the people that exists at I School is amazing. It makes someone feel they belong to be here.

Learning and applying new skills onto real world situations/ meeting everyone else and hearing their experiences

What You Hope to Achieve Here

Make friends and network

Enhance both technical and interpersonal skills

Broaden my skillset within information science

Leave with a UX research position

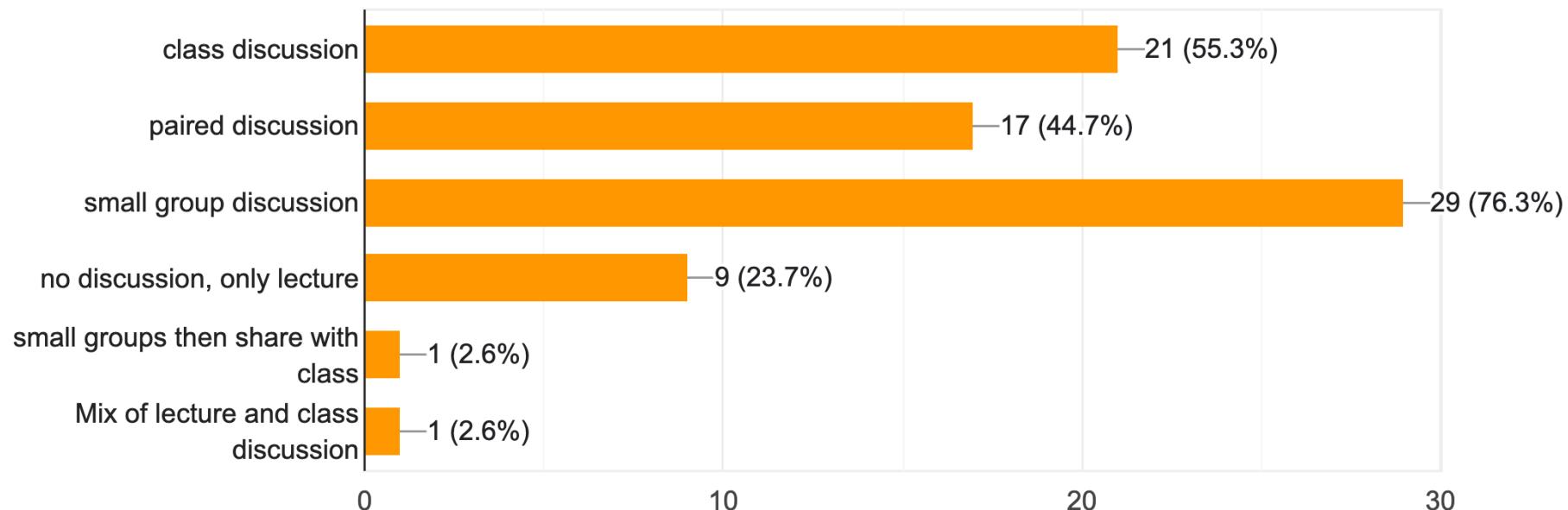
Get more technical in building software & upskill in design

Learn new skills in different areas & have fun

Be a sponge

What kinds of collaboration style do you like in class?

38 responses



Tools for Class Interaction



bcourses



#info-202-2025-fall



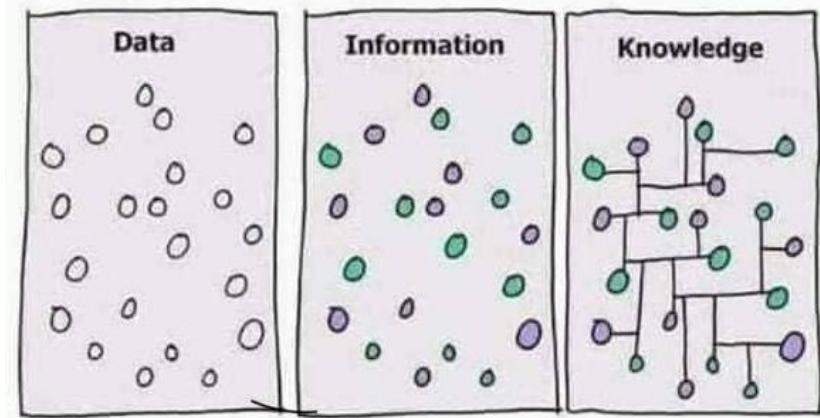
What is Data? What is Information?

Imagine you are viewing this boat from the dock.

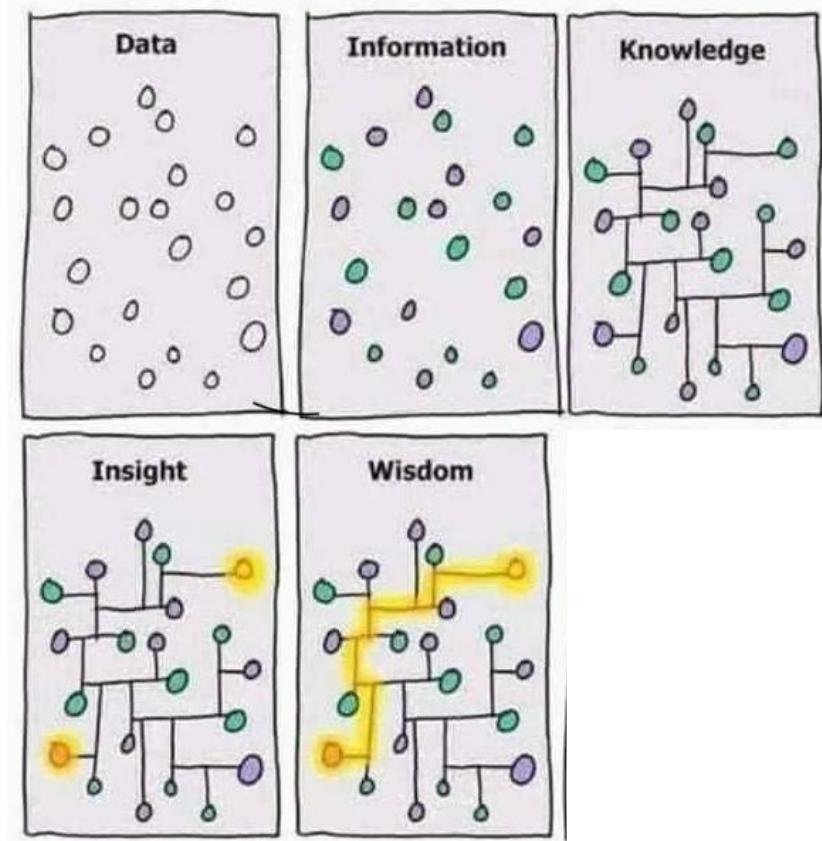


What kinds of **information** might you want to find out about it?

What kinds of **data** might you want to find out about it?



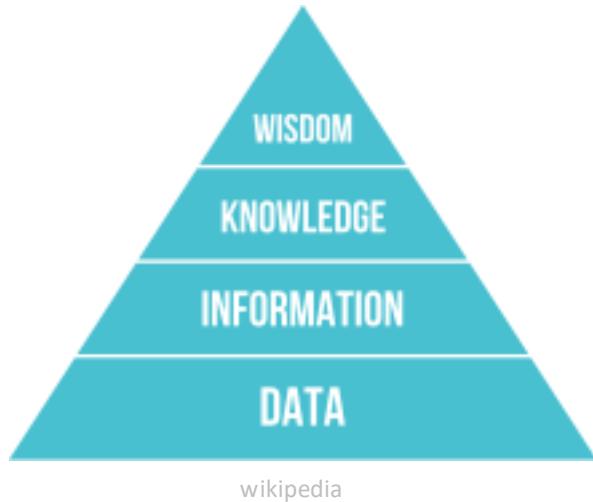
Original Image by Hugh MacLeod. <https://informationversusknowledge-blog.tumblr.com/>



Original Image by Hugh MacLeod. <https://informationversusknowledge-blog.tumblr.com/>

Data, Information, Knowledge, Wisdom

DIKW Pyramid



Data

The raw material of information

Information

Data organized and presented by someone

Knowledge

Information read, heard or seen and understood

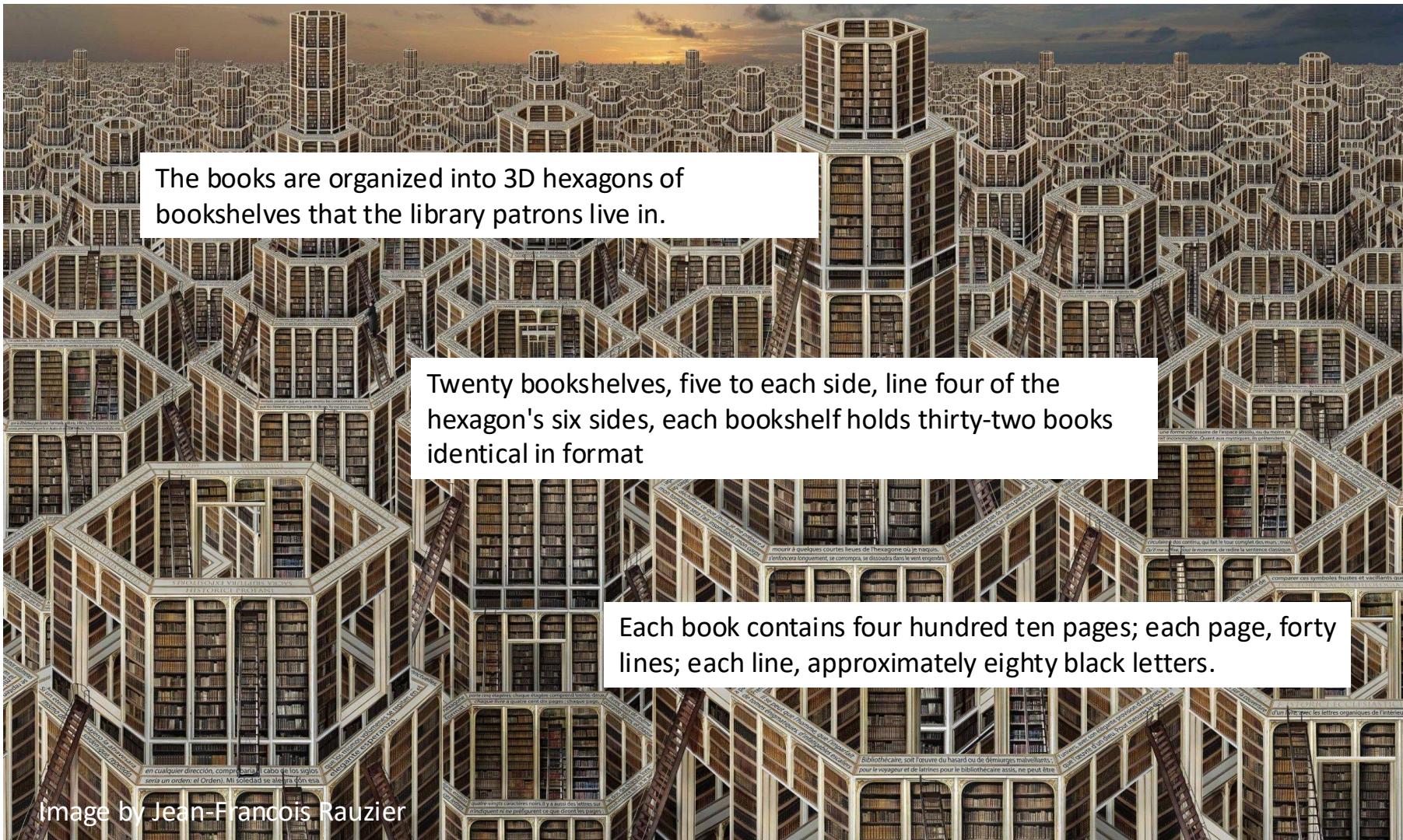
Wisdom

Distilled and integrated knowledge and understanding

People debate this representation; see the DIKW Wikipedia article

BORGES' LIBRARY OF BABEL

Imagine a universal library, containing books with every possible combination of 410 pages of letters, thus containing every book that ever has been and every book that ever could be written



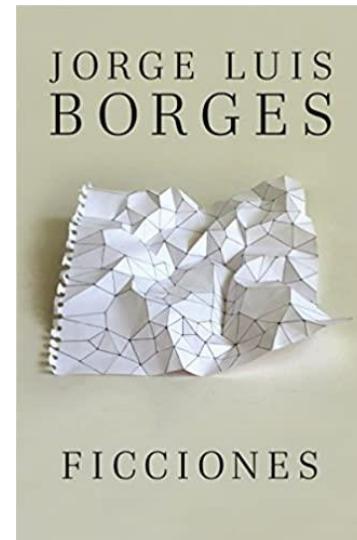
The books are organized into 3D hexagons of bookshelves that the library patrons live in.

Twenty bookshelves, five to each side, line four of the hexagon's six sides, each bookshelf holds thirty-two books identical in format

Each book contains four hundred ten pages; each page, forty lines; each line, approximately eighty black letters.

THE STRUCTURE OF THE BORGES LIBRARY

- The books are all 410 pages long, 40 lines per page, 80 characters per line, so 1,312,000 characters per book
- The alphabet has 25 characters
- The library contains **every possible** 410-page permutation of those characters, so 25 raised to 1,312,000 books.
- The books are organized into 3D hexagons of bookshelves that the library patrons live in.
- The entire library is a sphere; each hexagon is a center.



Vintage Español

THE DIGITAL BORGES LIBRARY

libraryofbabel.info by Jonathan Basile

A digital library of every possible 3200 character “book”

An Abstract from a New Scientific Paper

protein structures can provide invaluable information, both for reasoning about biological processes and for enabling interventions such as structure based drug development or targeted mutagenesis. after decades of effort, of the total residues in

exact match:

Title: **kcxrciiv.ulihymrgx** Page: 281

Location: [1hiffg50tskfegfqbcah52lm9f9gzs...-w4-s4-v19](#)

[more exact matches](#)

protein signatures can provide invaluable information, both for reasoning about biological processes and for enabling interventions such as structure based drug development or targeted mutagenesis. after decades of effort, of the total residues in human protein sequences are covered by an experimentally determined structure

An Abstract from a New Scientific Paper

Page 281 of 410

1hiffg50tskfe
gfqbcah52lm9f
9gzsmt2bwv8fx
...-w4-s4-v19

[Single Page](#)

[Anglighize](#)

[Bookmarkable](#)

[Download](#)

[Back to Portal](#)

protein structures can provide invaluable information, both for reasoning about biological processes and for enabling interventions such as structure based drug development or targeted mutagenesis. after decades of effort, of the total residues in human protein sequences are covered by an experimentally determined structure

THE DIGITAL BORGES LIBRARY

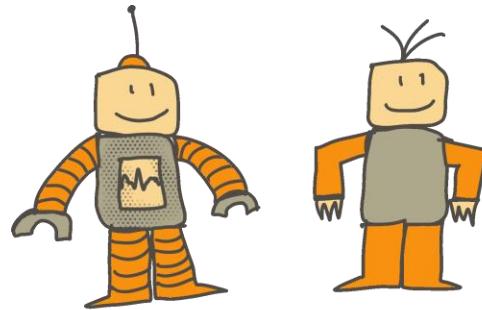
libraryofbabel.info by Jonathan Basile

A digital library of every possible 3200 character “book”

Exercise: Take a paragraph that you have written in the past. Will you be able to find it in the library? Why or why not? Try it out! What does it all mean?

ABOUT THIS COURSE

This course studies both the human and the technical sides of Info Org and Retrieval



and their intersection...



SCHOOL OF
INFORMATION

This course also
introduces you to many
of the instructors and
courses at the I School

ASSIGNMENTS, MIDTERM, PROJECT

Assignments help you practice concepts

Both during and after class

Some will build towards a larger assignment

We will teach you the concepts you need to use

Midterm will help you practice concepts; open book / take-home

Final project will let you explore a topic of your choice in depth

READINGS

When to read them?

People's preferences differ

I've posted them in advance

Assignments will *usually* be due the following Monday

READING REFLECTION ASSIGNMENTS



You'll be organized into groups of students



By the first deadline:

Read the readings
Post a thoughtful response to the prompt



By the second deadline:

Respond to other students' posts in a threaded conversation



Most interesting posts may be shared with full class

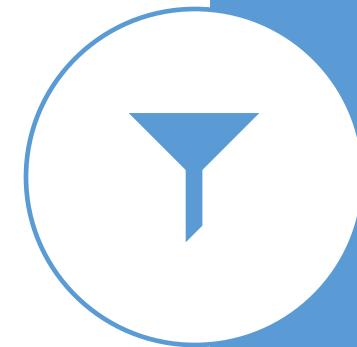
OTHER ASSIGNMENTS

- Create metadata structures (HTML, XML, JSON)
- Analyze a dataset
- Experiment with ranking & classification algorithms



A MINI-PROJECT

- Do grounded coding, individually and in pairs
- Build faceted category structures; assign to data
- Create faceted search interfaces from this data



Marti Co-Taught I 202 from 1997-2001, Solo in 2020-2021

www2.sims.berkeley.edu/courses/is202/f97/

EZproxy UCB lux-org/lux: Python... 1804.08199.pdf

**Information Systems
202**

**Information
Organization and
Retrieval**

[Description](#)

[Mission statement](#)

[Concepts](#)

[Outline](#)

[Instructors](#)

[Schedule](#)

[Enrolling](#)

[Grading](#)

[Announcements](#)

SIMS 202 schedule

This is a preliminary outline for the course. It can be expected to change during the semester.

Week, dates: Content

1. Aug 26 & 28: Introduction

- What is Information?
- How much is there in the world?
- What kinds of forms does it take?
- Why do people worry about information overload?
- Data, MetaData, The Information Lifecycle

[MAH Lecture 1 Slides](#)

[References used in Lecture 1](#)

[RRL Lecture 2 Slides](#)

2. Sept 2 & 4: Data Structures, Metadata, Data Models, Types of Data

[RRL Lecture 3 Slides](#)

[RRL Lecture 4 Slides](#)

[Library of Congress MARC Information](#)

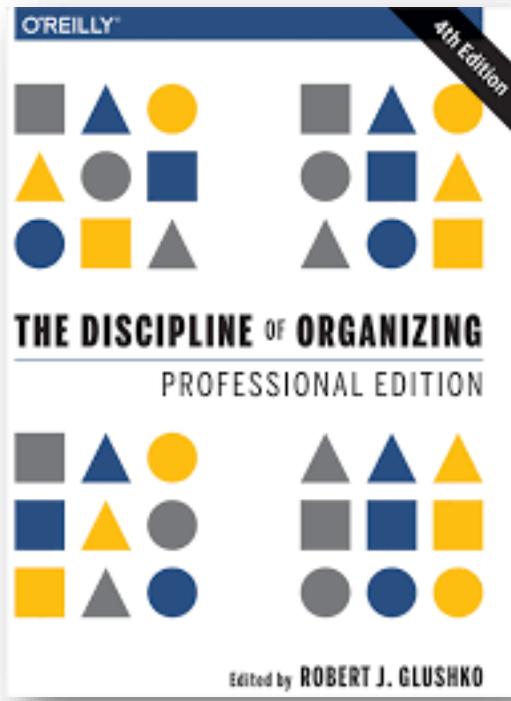


Prof. Ray Larson

Other Past Course Instructors



Prof. Bob Glushko



Prof. David Bamman

ANY
QUESTIONS?

