MAT 259 Visualizing Information

Organization & Classification of Data

Lecture 2, January 24, 2007

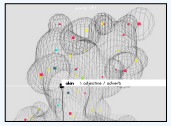
Data

- **Topic Discussion**
  - The definition of data
  - The classification and organization of data
- Data is the result of research, collection, discovery, creation (Shedroff)
- Data in itself not very useful: Its value is in how it is organized, transformed, and presented to give it meaning
- Context determines meaning for data

Two Forms of Data

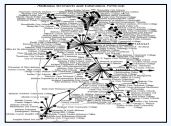
- **Entities:**
  - The objects we wish to visualize
- **Relationships:**
  - Define the structures and patterns that relate entities to each other
- **Sometimes relationships are explicit, other times, discovering the relationship becomes the purpose of visualization**

Types of Data



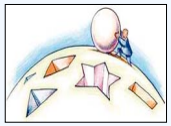
- **Datatypes (Spence)**
  - Numeric
  - Ordinal: Days of the week (ranking)
  - Categorical: Names of animals
- **Attribute Quality (Ware)**
  - Nominal: Labeling: (bus number, fruit, etc.)
  - Ordinal: Ranking
  - Interval: Relative difference (train sched)
  - Ratio: Object A is twice B

From Data to Information



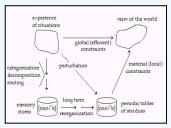
- 1st step: explore its organization
- Organization affects **interpretation** and **understanding**
- **Variations** in the organization of the same data set express **different** attributes and messages

Classification Methods (Shedroff)



- **Alphabetical**: arbitrary learned system
- **Numeric**: arbitrary learned system
- **Scalar**: (hotel star system) implies value scale
- **Sequential** (time): based on units
- **Spatial**: "sense of place"
- **Categories**: similar things grouped together
- **Associative:** (If a to b, then c to d)
- **Metaphoric**: A way to establish context
- **Random**: Creates complexity (game beginnings)

Systems of Classification

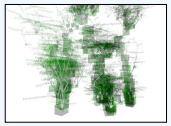


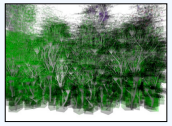
- We organize according to rules, systems (Linnaeus)
- but also according to experience (associative): Classification at level of **sensible properties** is a step towards rational ordering (Levi-Strauss)
- Classification based on **aesthetics**: weight, color, taste, shape, etc. even tough there may not be a necessary connection
- To classify is to speculate: a creative process

Explorative & Innovation

- To see same data sets in different organizations reveals **unexpected patterns** in the relationship of things
- To invent new forms of organization based on personal, or idiosynchratic rules enhances **novelty of experience**
- Nonetheless there needs to be some cultural **common ground**
- **Invention** is always a conversation with the conventional

Aesthetic & Informational Examples



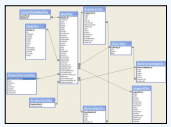


- **Associative**: Lisa Jevbratt 1:1
- **Biographical**: Daniel Spoerri's Anecdoted Topography of Chance
- **Affect**: Melanie Wein's http://www.the-fleetingness-of-bits.de/
- **Mapping:** Baby Name Voyager, Martin Wattenberg

Database/Data Structures

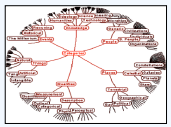
- A Database is an organized collection of data
- A collection of records stored in a systematic way
- Each record, a set of **data elements**, (basic unit of data such as name, street address, city, zip)
- Retrieval through any of the data elements
- **Relational model**: all data represented in terms of mathematical relationships

Metadata



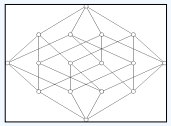
- Data about data: Information that describes another set of data
- Examples: Library catalog card, address book, etc.
- Metadata is what allows the organization, storage, retrieval of data

Hierarchical Tree Structure Model



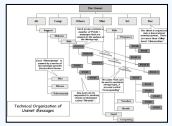
- Frequently hierarchical in structure, requiring parent/child relationship definitions
- Hard drive disk space
- Treemap (HCI, Maryland) History
- Internet, WWW
- Dewey Decimal System
- Cladistics: evolutionary relationships (cladograms)

Network Model



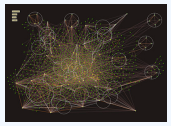
- Each record can have multiple parents and child records
- Organized in lattice structure consisting of links and nodes
- Lends easily to spatial visualization: Kohonen SOM map

Standards



- **Library of Congress**: **http://www.loc.gov/standards/**
- **Marc Standards**: **http://www.loc.gov/marc/**
- **Dublin Core**: a metadata standard for describing digital objects (including webpages) to enhance visibility, accessibility and interoperability, often encoded in XML
- **Harmony Project**: research methods and models for describing the variety of rich content **http://metadata.net/harmony/**

References



- Information Visualization, Robert Spence
- Information Visualization, Colin Ware
- **Information Visualization resources**, Stanford
- "**Metadating the Image**" Lev Manovich
- "**Readings in Information Visualization**" Ben Schneiderman, HCI, University of Maryland
- **Savage Mind** (Science of the Concrete), Claude Levi-Strauss
- "**Sorting Things Out, Classification & Its Consequences**" [Review] G. Bowker, S.Leigh Star