

# 5. Concepts and Categories

---

INFO 202 - 15 September 2008

Bob Glushko

## Plan for INFO 202 Lecture #5

---

The Vocabulary Problem

Varieties of Categorization Systems / Activities

Theories of Categorization and Meaning



# How Can We Ever Understand Each Other?

---

It sometimes seems a little amazing that people are able to communicate at all because they organize and name the world in such different ways

We establish a shared context by engaging in a dialogue

We use less frequent terms when we need to be precise

We can apply these ideas to the design of IR systems by making search iterative and by supporting complex query constructions

But the basic problems are deeply rooted in how we conceive of and categorize things and concepts

## What Are Categories?

---

Categories are *equivalence classes*

- sets of material and abstract things, processes, and events that we treat the same

This doesn't mean that every instance of a category is exactly the same but only that from some perspective or purpose we are treating them as equivalent based on something they have in common

Categories are cognitive / linguistic models for applying prior knowledge to new situations and recognizing instances or events

We use categories whenever we communicate, analyze, predict, or classify

# Varieties of Categorization Systems / Activities

---

Cultural Categorization

Individual Categorization ("Tagging")

Institutional Categorization ("Business Semantics")

## Cultural Categorization Systems

---

Embodied in culture and language

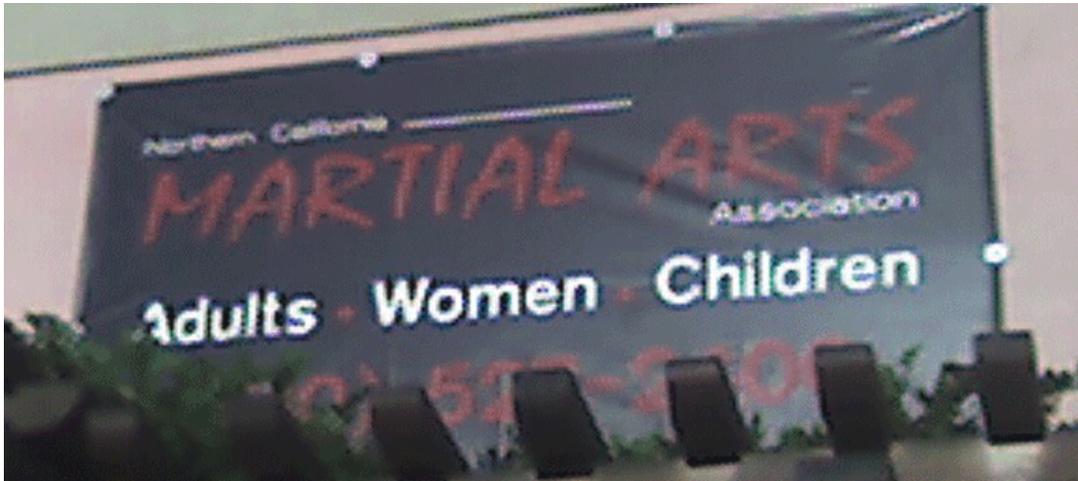
Acquired implicitly through development via parent-child interactions, language, and experience

Formal education can build on this, but non-formal cultural system can often dominate

Traditional perspective for thinking and research about categorization

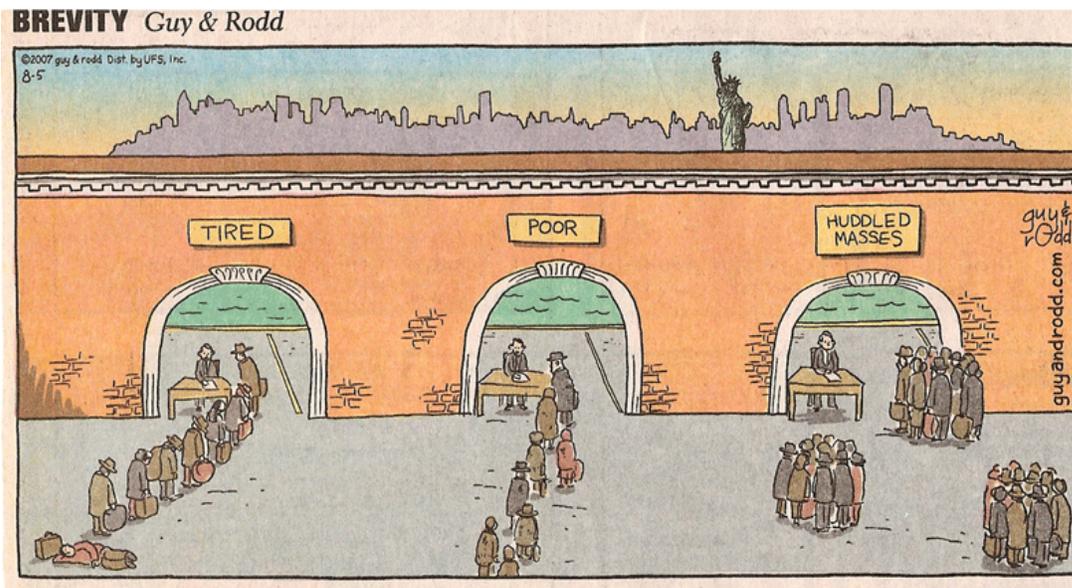
## Sexist Categorization in Martial Arts? (El Cerrito, CA 2007)

---



## Status of Liberty Categories ("The New Colossus," 1883)

---



# Individual Categorization Systems

---

A system developed by an individual for organizing a personal domain to aid memory, retrieval, or usage

Can serve social goals to convey information, develop a community, manage reputation

Have exploded with the advent of cyberspace, especially in applications based on "tagging"

An individual's system of tags in web applications is sometimes called a "folksonomy"

## What Should we Name this Store? (Ovando, MT 2007)

---



# Category Creativity {and,or,vs} Consistency?

---

In Flickr...

In del.icio.us...

## Institutional Categorization

---

Explicit construction of a semantic model of a domain to enable more control, robustness, and interoperability than is possible with just the cultural system

Often the collaborative artifact of many individuals who represent different organizational or business perspectives

Usually developed via rigorous and formal processes (e.g., in standards organizations)

Require ongoing governance and maintenance because of continuous changes taking place in related cultural and individual systems

# UN Standard Products and Services Codes for "Chicken"

---

**UNv91201**

Search Code:

Search Title:

Return  Records (Maximum 800 Records)

#	ID	Name
1	10101601	Live chickens
2	23220000	Chicken processing machinery and equipment
3	50111515	Chicken, minimally processed without additions
4	50111520	Chicken, minimally processed with additions
5	50112010	Chicken, processed without additions
6	50112011	Chicken, processed with additions

## Categories for CAFE

---



# "PT Helps DC Meet CAFE"

---

<http://www.edmunds.com/insideline/do/Drives/FirstDrives/articleId=43987>

The 2001 Chrysler PT Cruiser isn't what you surmise it to be. It's not a go-fast hot rod, it's not a mini-SUV, it's not a station wagon, and it's not a small van

The National Highway and Traffic Safety Administration (NHTSA) calls it a light truck

The Environmental Protection Agency (EPA) says it's a car

NHTSA overrules EPA, so PT helps DaimlerChrysler satisfy CAFE

# "Gross Income" Tax Code Categories

---

**(a) General definition**

Except as otherwise provided in this subtitle, gross income means all income from whatever source derived, including (but not limited to) the following items:

- (1) Compensation for services, including fees, commissions, fringe benefits, and similar items;
- (2) Gross income derived from business;
- (3) Gains derived from dealings in property;
- (4) Interest;
- (5) Rents;
- (6) Royalties;
- (7) Dividends;
- (8) Alimony and separate maintenance payments;
- (9) Annuities;
- (10) Income from life insurance and endowment contracts;
- (11) Pensions;
- (12) Income from discharge of indebtedness;
- (13) Distributive share of partnership gross income;
- (14) Income in respect of a decedent; and

# Principles of Categorization

---

*Similarity*

*Hierarchy*

Some levels in a category hierarchy seem more "natural" or "basic" than others

But categories can be nonhierarchical and overlapping

But we can readily create completely novel categories

So most things can be categorized in a very large number of ways and ANY characteristic of an object can be used to do so

## Borges' Categorization of Animals

---

(In a 1952 short story titled "The Analytical Language of John Wilkins" Borges writes:)

in a certain Chinese encyclopedia "it is written that 'animals are divided into:

(a) belonging to the Emperor, (b) embalmed, (c) tame, (d) suckling pigs, (e) sirens, (f) fabulous, (g) stray dogs, (h) included in the present classification, (i) frenzied, (j) innumerable, (k) drawn with a very fine camelhair brush, (l) et cetera, (m) having just broken the water pitcher, (n) that from a long way off look like flies."

# Yahoo's Categorization of Animals

---



## Categories Evolve

---

### Animals by Alphabetical and Taxonomic Groups

- [Complete List of Animals by Name](#) (588)
- [Arachnids@](#)
- [Arthropods](#) (875)
- [Birds](#) (1082)
- [Bryozoans](#) (16)
- [Cnidarians](#) (46)
- [Crustaceans@](#)
- [Echinoderms](#) (16)
- [Fish](#) (605)
- [Insects@](#)
- [Mammals](#) (4131)
- [Mollusks](#) (112)
- [Reptiles and Amphibians](#) (519)
- [Sponges](#) (3)
- [Worms](#) (55)

# Properties / Features for Instances and Categories

---

It seems intuitive to think in terms of properties or features when describing instances and categories as a way to understand what instances have in common

Some features may be absolutely necessary for every instance of the category

Some features are *additive*

– they can be added to the set of features an instance already has

Other features are *substitutive*

– once added, they preclude other features on the same dimension

# Lakoff - "Women, Fire, and Dangerous Things"

---

Professor at UC Berkeley (Department of Linguistics)

Very influential figure in thinking about human categorization, and more recently on metaphor and politics of language use

WFDT provides summary of historical work and proposes a model of cognition and categorization

Contains clear explanations using examples

# Is Meaning Objective or Experiential?

---

Lakoff contrasts two fundamental theories about how we experience and understand the world

- *OBJECTIVIST or TRANSCENDENTAL or CLASSICAL perspective:*  
Meaningful thought and reason involve abstract symbol manipulation and their correspondence to an objective reality, independent of the organism doing the thinking
- *EXPERIENTIAL perspective:*  
Meaningful thought and reason reflect the nature of the organism doing the thinking - including the nature of its body, its interaction in its environment, its social character, and so on

Both perspectives treat CATEGORIZATION as the main way that we make sense of experience, but they differ widely in how they explain how categories are defined and used

---

## The Objectivist View

---

The mind is an abstract machine

Thought is mechanical manipulation of symbols

Symbols get their meaning from correspondences to the external world

Thought can be broken down into simple "building blocks"

How thoughts combine is defined by mathematical logic

# The Objectivist View of Categories

---

Categories have clear boundaries defined by necessary and sufficient properties

- *Necessary*  
means that every instance must have the property to be in the category
- *Sufficient*  
means that any instance that has the necessary properties is in the category

Thus all members of the category have equal status in the equivalence class

Example: A prime number is an integer divisible only by itself and 1

## Frege's Objectivist Philosophy of Language [1]

---

Gottlob Frege (1848-1925), father of modern logic, investigated the idea of equivalence – how can you tell that two things are the same – in "Über Sinn und Bedeutung"

"Sinn" or "sense" (or "intension") – the inner concept that people understand; words have intensions

"Bedeutung" or "reference" (or "extension") – the thing being referred to; the set of all objects in the world that can be described by the concept; intensions belong to extensions

# Frege's Objectivist Philosophy of Language

## [2]

---

According to Frege, you understand the intension by looking up its extension ("in the world")

You can tell that "Bob" and "Matt" mean different things because the people who answer to the names are different

"Bob" and "Dr. Bob" must mean the same thing because I respond to both of these names

Some categories can be defined by *ENUMERATION* - listing all of the instances

---

## If Names Mean What They Refer To...

---

Prune == Dried Plum

Chinese Gooseberry == Kiwi Fruit

Patagonian Toothfish == Antarctic Cod == Chilean Sea Bass

Sectarian Conflict == Civil War

Terrorists == Insurgents == Resistance == Freedom Fighters

# Wittgenstein

---

Ludwig Wittgenstein (1889-1951) – "philosophy of ordinary language" - first to discuss problems with classical theory

Dismantles Frege in "Philosophische Untersuchungen"

Agrees with Frege that where the extensions have fixed characteristics or can be enumerated you can understand words by following the association to their extensions

But rebuts Frege with argument that there are no fixed extensions for most words

## Wittgenstein's Rebuttal to Frege: Meaning is Use

---

There may be defining features for typical instances

But there are no features that are necessary and sufficient for all examples of the category

Even when features can be identified, they change in different contexts and over time

Different instances vary substantially in how typical or representative they are of the category even though they share all the required features

# Necessary and Sufficient Properties - Not!

---

Wittgenstein's Counterexample is "Game"

- No common properties are shared by all games
- Some involve competition, others are cooperative
- Some involve physical skill, others more mental skills, others luck
- Some require equipment, others don't
- Some involve teams, others are solitary

No fixed category boundary - we can extend game category to include video games, networked games, etc.

# Family Resemblances

---



Instances of a category often share many features, but some instances might have properties that are not widely shared

These widely shared though not universal properties produce *FAMILY RESEMBLANCES* among the members

# Characteristic Features

---

Perceived degree of category membership has to do with which features help define the category

Members usually do not have ALL the necessary features, but have some subset

Those members that have more of the central features are seen as more central members

# Gradience in Category Membership

---



Not all members of a category are equally good examples

The perceived centrality or typicality of category membership depends on the extent to which the most characteristic properties are shared

"Someone says to me 'Show the children a game.' I teach them gaming with dice, and the other says 'I didn't mean that kind of game.' "  
(Wittgenstein)

# The Experiential View of Thought and Categorization

---

Thought is embodied; our conceptual systems are grounded in perception, our body motion, and our physical and social experiences

Thought is imaginative

Thought has gestalt properties and cannot be reduced to building blocks or combinations of them

Conceptual structure can be described using cognitive models that have the above properties

## Category Hierarchies

---

Categories can be organized into a hierarchy from the most general to the most specific

Instances are multiply categorized: my pet cat can be described as

- An animal
- A mammal
- A cat
- An American Shorthair
- Boris

# Basic-Level Categories

---

In the middle of category hierarchies are those that more "basic" because the within-category differences are smallest and the between-category differences are the largest

This means that perceptual, cognitive, and motor functions are "sharpest" at this level at identifying and thinking about category membership

Compare "Let the mammal out" or "Let the American Shorthair out" vs "Let the cat out" or "Let Boris out"

## Summary: Why Study Categorization?

---

Categorization is central to how we organize information and the world, and categories are involved whenever we communicate, analyze, predict, or classify

- Informally with "cultural" categories
- Formally with "institutional" categories
- Whenever we design data structures, programming language class hierarchies, user or application interfaces, ...

Categorization is much messier than our computer systems and applications would like

But understanding how people (and each of us) categorize can help us design better systems and interfaces

## Readings for INFO 202 Lecture #6

---

Svenonius Chapter 3, Chapter 4 (62 - 66)

Lois Chan and Marcia Zeng, "Metadata Interoperability and Standardization: A Study of Methodology Part I. Achieving Interoperability at the Schema Level", D-Lib Magazine, 12(6), June 2006

Cory Doctorow, "Metacrap: Putting the torch to seven straw-men of the meta-utopia"