I 202: INFORMATION ORGANIZATION & RETRIEVAL
FALL 2025

Class 12: Lexical Relations

MORE RECENT HTML DEVELOPMENTS

- HTML5 includes some tags that assign meaning to the

  structural elements
    - *<nav>, <article>, <section>, <aside>, etc*

- Data structuring standards are growing in importance
    - *JSON-LD, RDFa, making us of Schema.org markup*

- Purposes:
    - *Aid in machine parsing, for search engine optimization*
    - *Align with accessibility goals to aid screen readers*

Non-Semantic HTML Code

```html
<!DOCTYPE html>
<body>
    <div>
        <span>My Website</span>
        <div>
            <a href="#home">Home</a>
            <a href="#about">About</a>
            <a href="#contact">Contact</a>
        </div>
    </div>
    <div>
        <div>
            <div>Welcome to My Website</div>
        </div>
        <div>
            <div>This is the main content of my website.</div>
        </div>
    </div>
    <div>
        <span>&copy; 2025 My Website</span>
    </div>
</body>
```
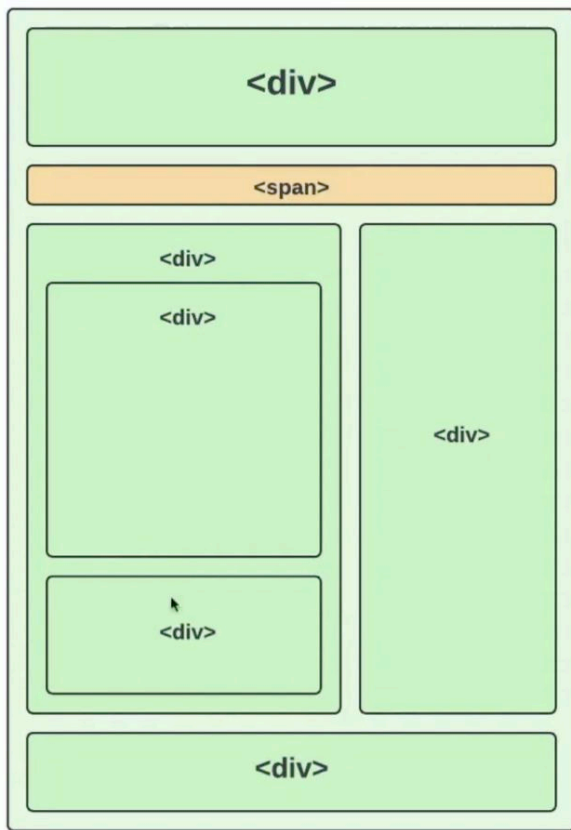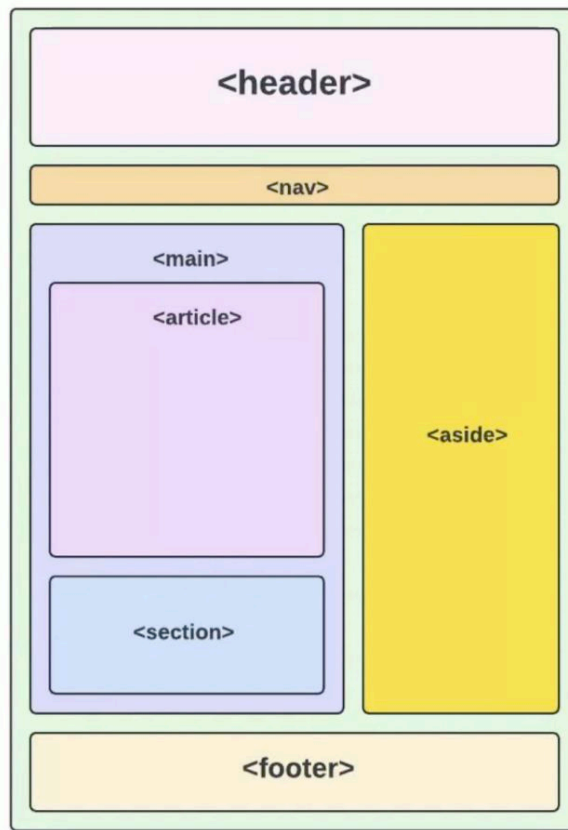
Semantic HTML Code

```html
<header>
    <h1>My Website</h1>
    <nav>
        <ul>
            <li><a href="#">Home</a></li>
            <li><a href="#">About</a></li>
            <li><a href="#">Contact</a></li>
        </ul>
    </nav>
</header>

<main>
    <h2>Welcome to My Website</h2>
    <p>This is the main content of my website.</p>
</main>

<footer>
    <p>&copy; 2025 My Website</p>
</footer>
```

semrush.com

SEMRUSH

https://www.semrush.com/blog/semantic-html5-guide/

Suk Min Huang's lab notes: https://groups.ischool.berkeley.edu/i253/fl25/slides/week1-lab1.pdf

Semantic HTML tags

Tag	Description
<article>	Defines independent, self-contained content
<aside>	Defines content aside from the page content
<details>	Defines additional details that the user can view or hide
<figcaption>	Defines a caption for a <figure> element
<figure>	Specifies self-contained content, like illustrations, diagrams, photos, code listings, etc.
<footer>	Defines a footer for a document or section
<header>	Specifies a header for a document or section
<main>	Specifies the main content of a document
<mark>	Defines marked/highlighted text
<nav>	Defines navigation links
<section>	Defines a section in a document
<summary>	Defines a visible heading for a <details> element
<time>	Defines a date/time

Pages with element or role
Web Almanac 2024: Accessibility

■ footer ■ header ■ main ■ nav

Year
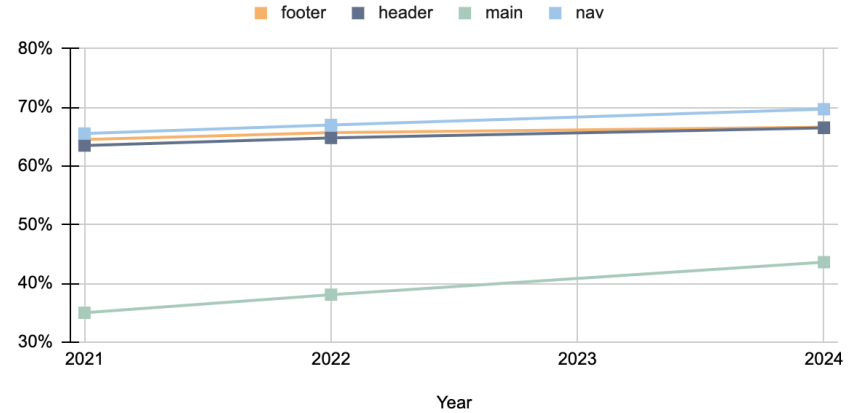
This Week's Focus

Data / Information
Collections
Categories
 - Types of categories
 - Cognitive / language aspects
        - **Naming / Lexical similarity**
 - Structure
        - Hierarchical / Taxonomy
        - Faceted
        - Overlapping / Clustering
        - Network / Ontology
 - Use in Navigation & Search
        - **Information Architecture**
        - Faceted Navigation

Technology Support for Info Org
        - Identifiers
        - Metadata
        - Markup
        - Schema / Databases
        - Search Ranking / Evaluation
        - Automated category creation
        - **Automated similarity**

Social / Ethical Aspects
        - Cultural Bias
        - Intellectual Property
        - Standards Process

Today's Outline

The Vocabulary Problem

Semantic vs Lexical

Lexical Relations

WordNet

Controlled Vocabs in IA

THE VOCABULARY PROBLEM

The same idea can be (and is) expressed an astonishing number of ways.

This variability has its roots in how we think and use language.

This causes problems for many IT systems.

From Lecture 3: Representations of Names Can Be Noisy

Figure 1: Top 25 Representations of "Knockin' On Heaven's Door" [35]

```
Guns N' Roses    -   Knockin' On Heaven's Door
Guns N' Roses    -   Knocking On Heavens Door
Guns 'N' Roses   -   Knockin' On Heaven's Door
Guns N' Roses    -   Knockin On Heavens Door
Guns N' Roses    -   Knockin' On Heavens Door
Guns N'roses     -   knockin on heavens door
Guns N' Roses    -   Knocking On Heaven's Door
Guns N Roses     -   Knockin' On Heaven's Door
Guns N Roses     -   Knockin On Heavens Door
Guns And Roses   -   Knocking On Heavens Door
Guns Nroses      -   Knockin On Heavens Door
Guns'n' Roses    -   Knockin' On Heaven's Door
Guns N Roses     -   Knocking On Heavens Door
Guns'n'Roses     -   Knockin' On Heaven's Door
Guns 'N Roses    -   Knockin' On Heaven's Door
Guns & Roses     -   Knockin' on Heaven's Door
Guns N'roses     -   Knockin' On Heaven's Door
Guns and Roses   -   Knockin' On Heaven's Door
Guns'n Roses     -   Knocking On Heavens Door
Guns 'n' Roses   -   Knockin' On Heavens Door
Aerosmith        -   Knocking On Heaven's Door
Guns 'n Roses    -   Knocking On Heaven's Door
Guns 'n' Roses   -   Knocking On Heavens Door
Guns N Roses     -   Knocking On Heaven's Door
Guns N' Roses    -   Knockin On Heaven's Door
```

Jones, R. Fingerprinting and Metadata Progress Report. Last.fm: The Blog, 2008., via Hemerly, Making Metadata: The Case of MusicBrainz, SSRN 2011

Vocabulary Problem Example: Naming for Smart Rooms

EXERCISE:

What should the names of devices be to allow for controlling each component by voice?

pollev.com/I202

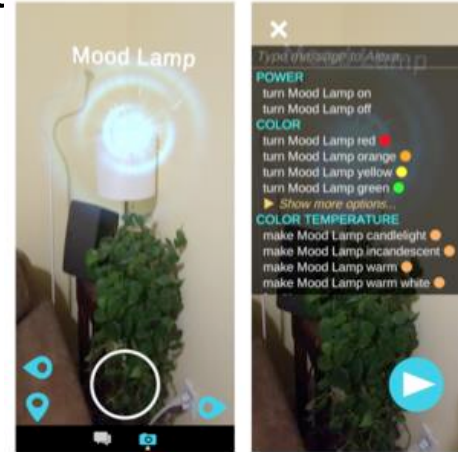Meghan Clark, One-Shot Interactions with Intelligent Assistants in Unfamiliar Smart Spaces, UCB Dissertation, 2021

VOCABULARY PROBLEM EXAMPLE: NAMING FOR SMART ROOMS

How did she solve this?

Augmented reality and autosuggest!

Meghan Clark, One-Shot Interactions with Intelligent Assistants in Unfamiliar Smart Spaces, UCB Dissertation, 2021

Summary: Vocabulary Problem Example: Naming for Smart Rooms

- **Device names were mostly not guessable**

- **Popular names could apply to multiple devices**

- **The order of naming had an effect**
  - *If someone started with the left blind group, they would call them "blinds", thus affecting subsequent names*

- **Names do not always capture relationships**
  - *Group: "spot lights" but left light called "recessed light"*

- **Frames of reference can differ**
  - *Front / back vs left / right – depended on assumptions of orientation*

LET'S DIVE DEEPER INTO LANGUAGE'S COMPLEXITIES

Semantic vs Lexical

Semantic vs Lexical

- **Semantic**: the **meaning**

- **Lexical**: how the meaning is **expressed** in words

Debby Hudson

"What's in a name? That which we call a rose
By any other name would smell as sweet"
Shakespeare, Romeo and Juliet

LEXICAL RELATIONSHIPS

- Synonymy

- Polysemy

- Homography

- Hyperonymy / Hyponymy

- Metonymy

Synonyms

- Different lexical forms that have the same meaning in some or all contexts

  More formally, two lexemes are strict synonyms if:
  - They can be substituted for each other in all situations
  - They have the same meaning

  couch / sofa
  auto / car

- Very few synonyms meet the strict requirement:

  This is a big day.
  🚫 This is a large day.

  Nurse, how did the patient do today? Doctor, he vomited this morning.
  🚫 Nurse, how did the patient do today? Doctor, he puked this morning.

POLYSEMY

Polysemous words have multiple senses -- often loosely related

"bank" (financial) has several related senses:

bank: *Let's walk over to the bank.* (meaning the building on Shattuck)

bank: *The bank's stock went up today.* (an institution, e.g., Wells Fargo)

bank: *Put your money in the bank.* (where money is kept, abstract)

HOMOGRAPHY VS. POLYSEMY

- **A homograph is a word with multiple senses that are not related**
  - *bat (stick), bat (flying mammal)*

- **Polysemy: senses that are closely related (close to the prototype)**
  - *money (a medium of exchange), money (as wealth; he comes from money)*

- **What looks like homography may be polysemy, often for historical reasons**
  - *mole (the animal), mole (the spy)*

- **Homophones: like homographs, but have different spellings but pronounced the same way**
  - *write / right*

HYPONYMY AND HYPERNYMY

- When words encode IS-A or inclusion relationships, the word for the more specific class is the HYPONYM and the other is the HYPERNYM

- This builds a "lexical hierarchy" that represents the "semantic hierarchy"

- Often used to situate basic categories with respect to superordinate and subordinate categories

vehicle is a hypernym of car
mango is a hyponym of fruit

Wikipedia
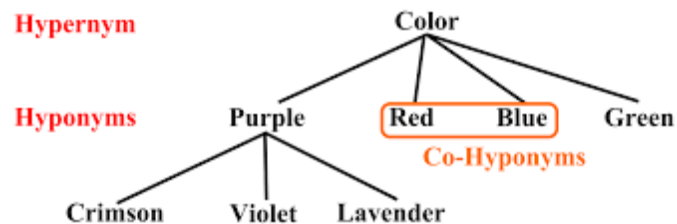
DIFFERENT LEXICAL FORM

SAME LEXICAL FORM

hypernyms
(superordinate)

polysemes

WORD

hyponyms
(subordinate)

homographs

DIFFERENT LEXICAL FORM

SAME LEXICAL FORM

hypernyms
(superordinate)

cooking utensil

polysemes

pot (betting)

pot (flower)

pot

hyponyms
(subordinate)

kettle

coffee pot

pot

homographs

DIFFERENT LEXICAL FORM	SAME LEXICAL FORM
hypernyms (superordinate)	hypernyms (superordinate)
synonyms	polysemes
sibling terms	homographs
hyponyms (subordinate)	hyponyms (subordinate)

WORD

hypernyms
(superordinate)

fundamental measure

hypernyms
(superordinate)

---

synonyms

time

polysemes

period (menstrual)

period
(amount of time)

sibling terms

temperature, mass, length

homographs

period (punctuation)

hyponyms
(subordinate) week, past, eve…

hyponyms
(subordinate)

trial period, test period

Meronymy and Holonymy
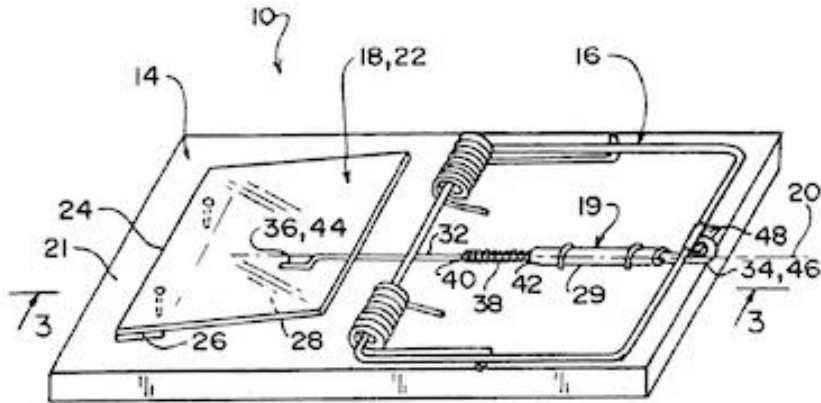
From U.S. Patent No. 6,655,077, titled "Trap for a mouse."

The lexical expression of the **part-whole relation.**

wire vs. trap:
Wire is **meronym** of trap
Trap is **holonym** of wire

Metonymy

Substituting the name of an attribute or feature for the name of the thing itself.

UC Berkeley

"We are hearing from California Hall that we have budget cuts."

"She has a good ear for music." (metonymy and polysemy)

THERE ARE MORE! LANGUAGE IS FUN!

David Bamman
I School Professor

- Applied Natural Language Processing (I 256)
  - *Oriented towards interdisciplinary students*
  - *Focus on using existing tools*

- Natural Language Processing (I 259)
  - *In-depth on algorithms and implementation*

WORDNET

- The big kahuna of lexical resources (>100,000 nouns)
  - *Nothing else was free and online when it started*

- Main concept: the **synset**

- **Instead of alphabetical ordering, organized by semantic properties and relationships**

- Includes useful(?) functions
  - *Similarity*
  - *Stemming*

- Created by psychologists & linguists

- Originally English only, now many languages
  - *http://compling.hss.ntu.edu.sg/omw/*

visuwords.com

Find synonyms, meronyms, antonyms, etc
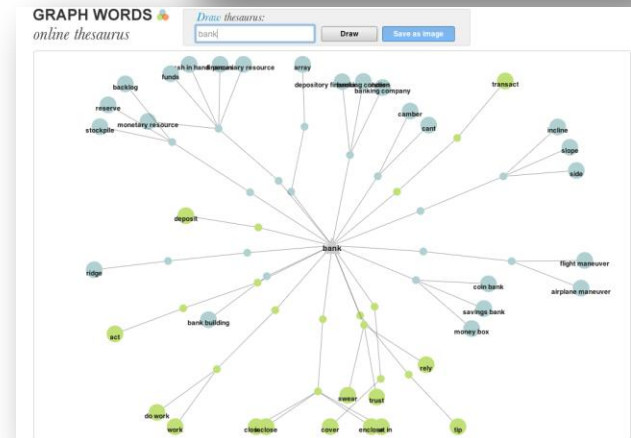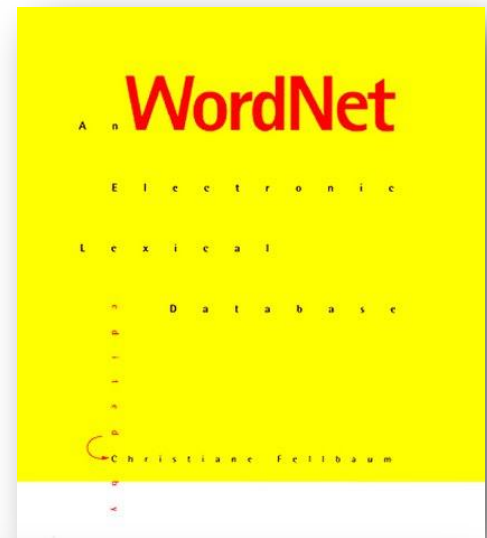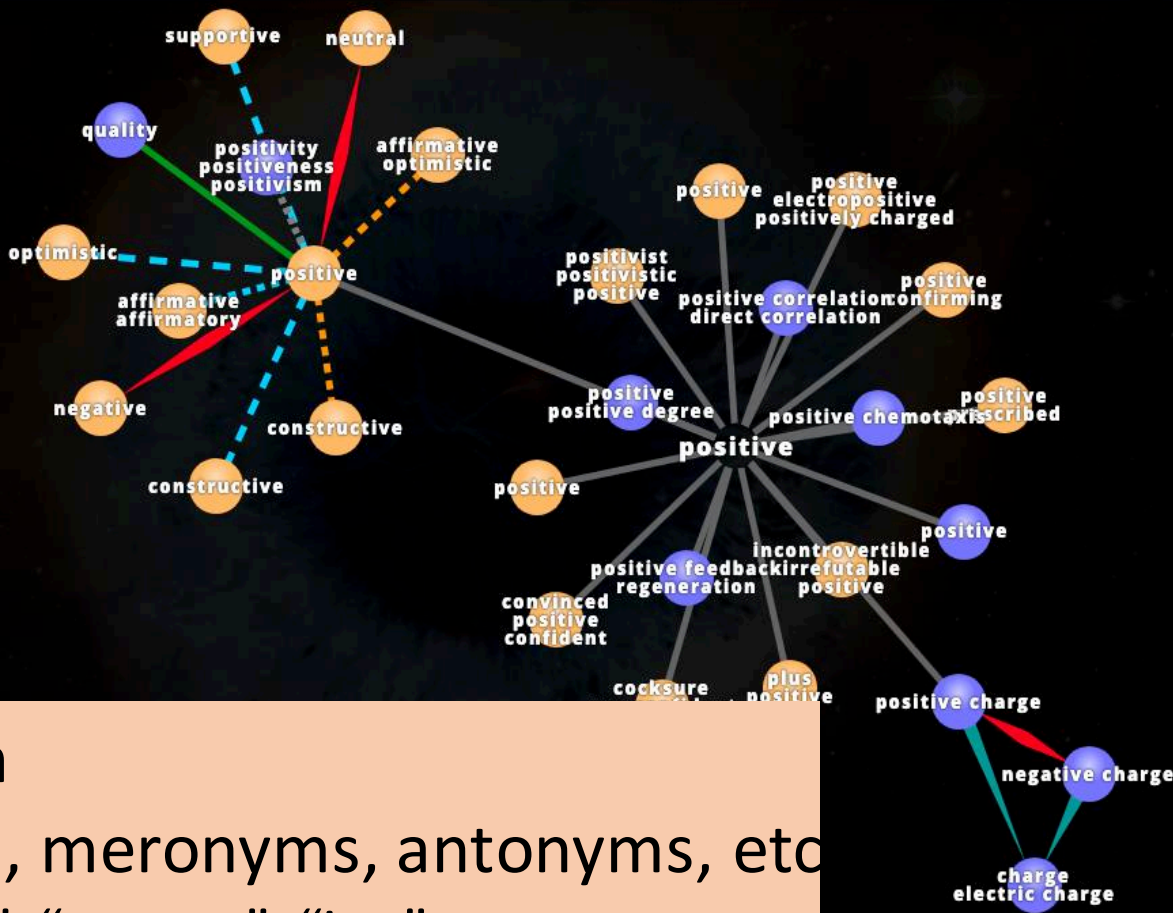
Examples: "bank", "money", "joy"

WHAT IS WRONG WITH WORDNET

Missing words like "kahuna"

Missing links

Missing many noun-noun compounds

Uneven hierarchy levels

General; often you need specific vocabularies

CONTROLLED VOCABULARIES IN INFORMATION ARCHITECTURE

O'REILLY®

4th Edition

Information
Architecture

FOR THE WEB AND BEYOND

Louis Rosenfeld,
Peter Morville & Jorge Arango

What to take from this reading

Ch 10:
Thesauri, Controlled Vocabularies, & Metadata

- Another definition of Metadata

- Synonym rings, authority files, and a different meaning of "thesauri"

- How these are used in websites

- (Only assigned to read through Thesaurus Standards, but feel free to read their take on Poly Hierarchy & Faceted Metadata)

Types of Controlled Vocabularies

(Vocabularies)

Synonym Rings	Authority Files	Classification Schemes	Thesauri

Simple ←———————————————————————————→ Complex

Equivalence	Hierarchical	Associative

(Relationships)
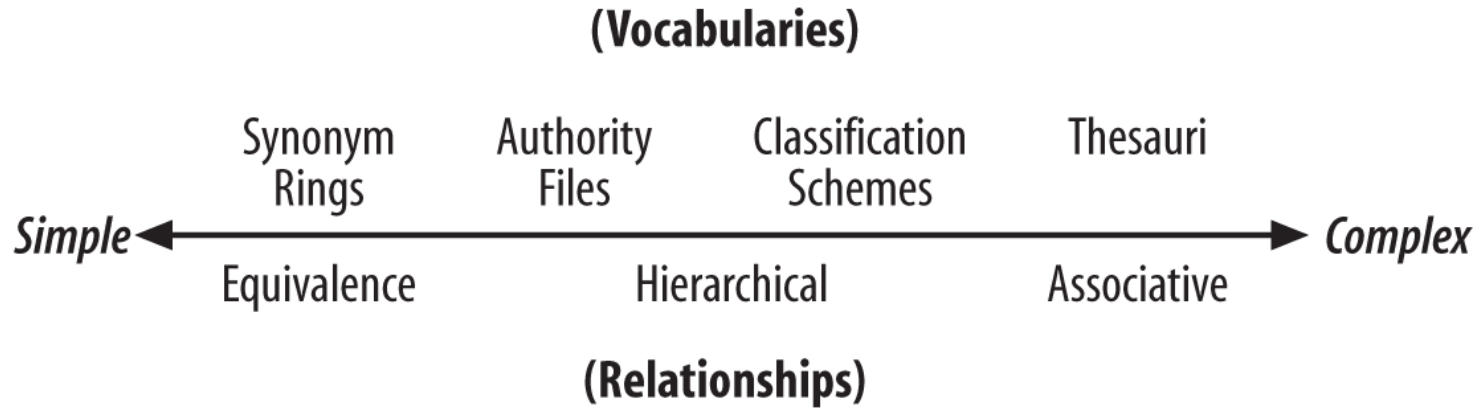
FROM LECTURE 3: CONTROLLED VOCABULARY:
TRY TO REDUCE THE NOISY NAME PROBLEM

"A fixed or closed set of description terms in some domain with precise definitions that is used instead of the vocabulary that people would otherwise use." -- TDO

FROM LECTURE 3:
TYPES OF CONTROLLED VOCABULARIES

- Dictionaries
- Authoritative names
- Identifier Standards
- Subject heading lists
- Classification systems

Today's reading / lecture adds a few more related ideas

"Synonym Ring"

Correspondence to lexical relations?
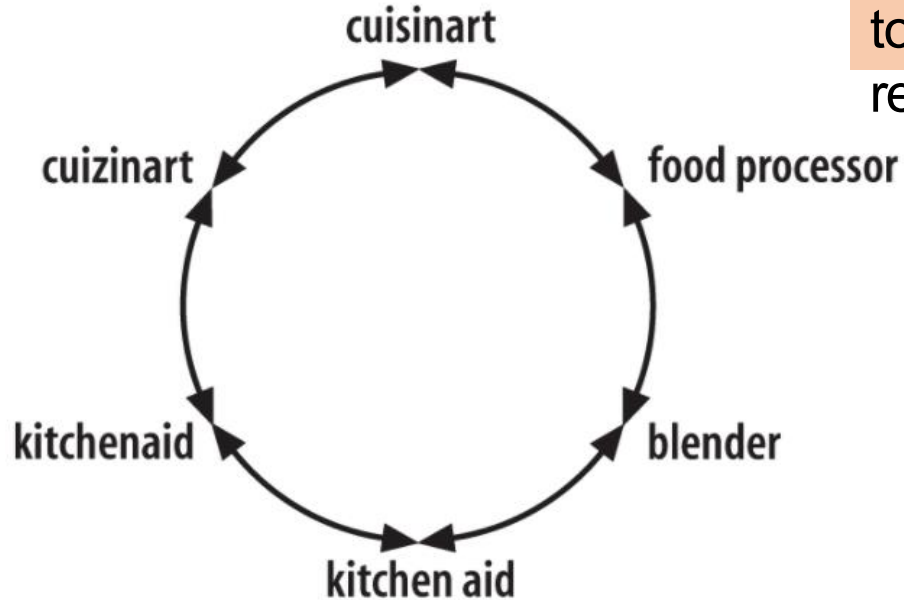
Lexical Relations, Pairwise on Arrows

orthographic variant

hypernym

sibling

sibling

orthographic variant

hyponym

cuisinart

food processor

cuizinart

blender

kitchenaid

kitchen aid
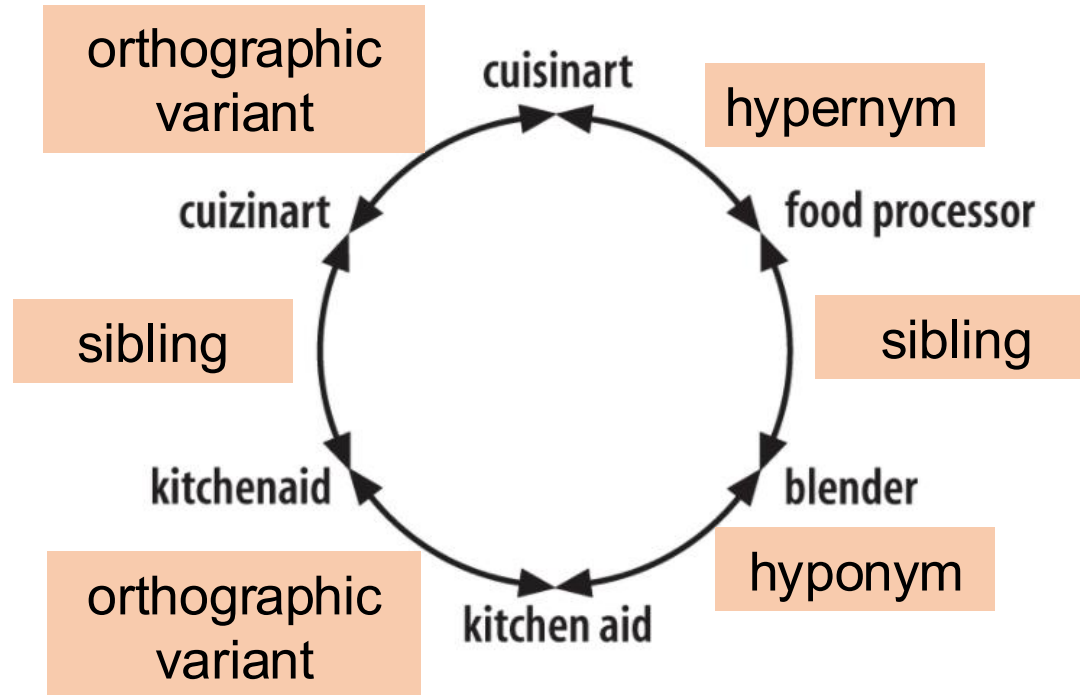
Authority Files

Indicates the preferred term(s) among a list of related terms

Example: correct spelling of a misspelled brand name

Rosenfeld et al., Information Architecture, O'Reilly, 2015

Authority Files

Indicates the preferred term(s) among a list of related terms

Example: scientific (or generic) name of a drug

Authority Files

Indicates the preferred term(s) among a list of related terms

Example: scientific (or generic) name of a drug

TYLENOL (ACETAMINOPHEN)
- TYLENOL W/ CODEINE (ACETAMINOPHEN; CODEINE PHOSPHATE)
- TYLENOL W/ CODEINE NO. 1 (ACETAMINOPHEN; CODEINE PHOSPHATE)
- TYLENOL W/ CODEINE NO. 2 (ACETAMINOPHEN; CODEINE PHOSPHATE)
- TYLENOL W/ CODEINE NO. 3 (ACETAMINOPHEN; CODEINE PHOSPHATE)

THESAURUS

- Sense 1: the familiar book we use to find alternative words when writing

- Sense 2: used for vocabulary management in information systems;
  - *Synonym management for an information system*
  - *Many kinds of relations (like ontologies have)*

Semantic Relationships in a Thesaurus
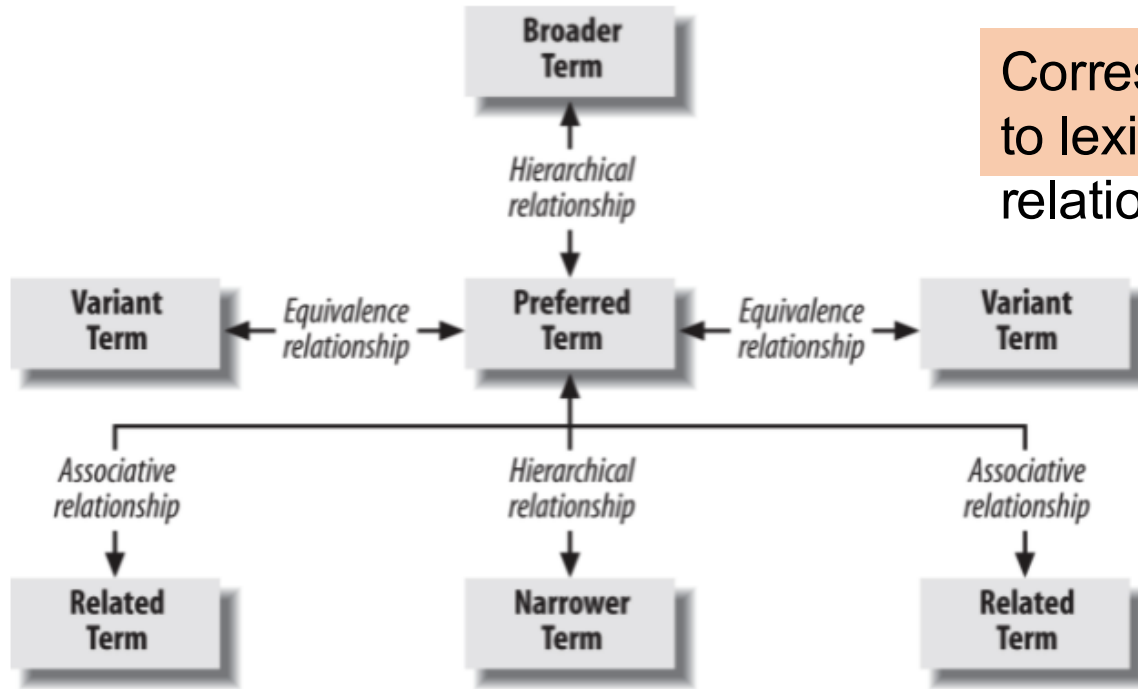
Figure 10-11. Semantic relationships in a thesaurus

Correspondence to lexical relations?

Semantic Relationships in a Thesaurus

hypernym

synonym

Broader Term

Hierarchical relationship

Variant Term ← *Equivalence relationship* → **Preferred Term** ← *Equivalence relationship* → **Variant Term**

Associative relationship

Hierarchical relationship

Associative relationship

Related Term

Narrower Term

Related Term

sibling / meronym / other relationship

hyponym

Figure 10-11. ... n a thesaurus

Semantic Relationships in a Thesaurus

Correspondence to ontologies?

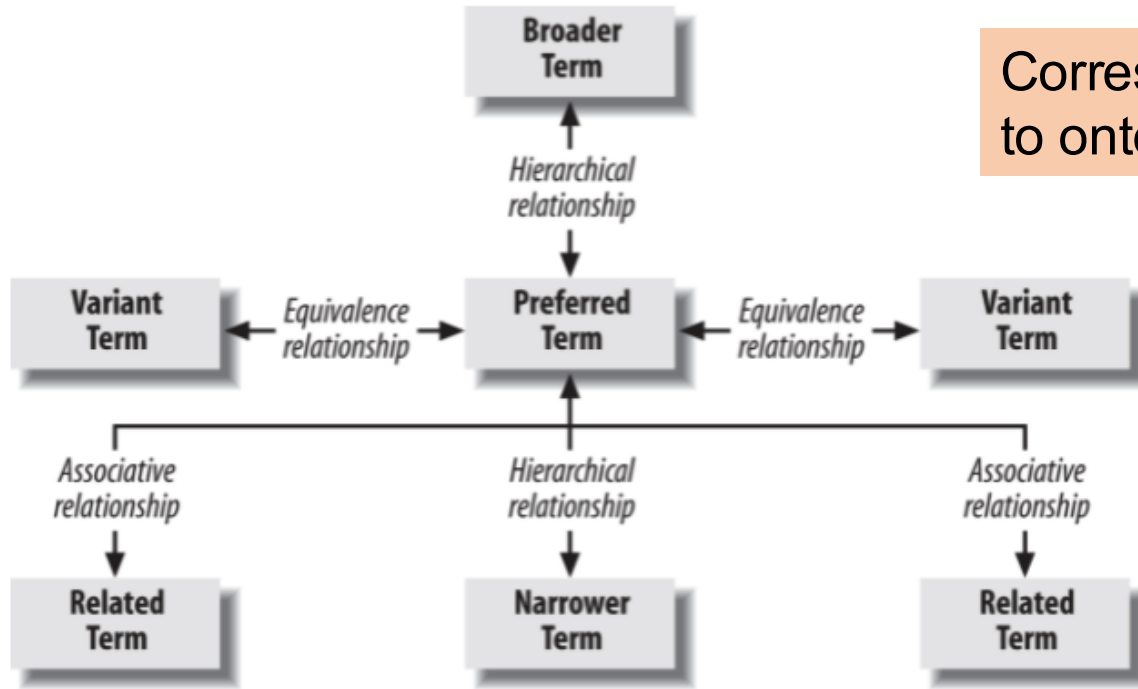Figure 10-11. Semantic relationships in a thesaurus

THESAURUS (FOR VOCABULARY MANAGEMENT)

- A THESAURUS in indexing is a tool for finding the "right" or "good" terms of a controlled vocabulary

- A collection of vocabulary terms annotated with lexical relationships:

  - Preferred (UF "used for")

  - Broader (BT "broader term")

  - Narrower (NT "narrower term")

  - Related (RT "related term" or "see also")

NEXT LECTURE

- Semantic Similarity

- Taxonomic vs Thematic Relations