Principles of Information and Database Management
198:336
Week 3 – Feb 7

Matthew Stone

Today

Conceptual modeling

- What concepts are there?
- How are concepts related?
- How do you find them?
- How do you represent them for DBs?

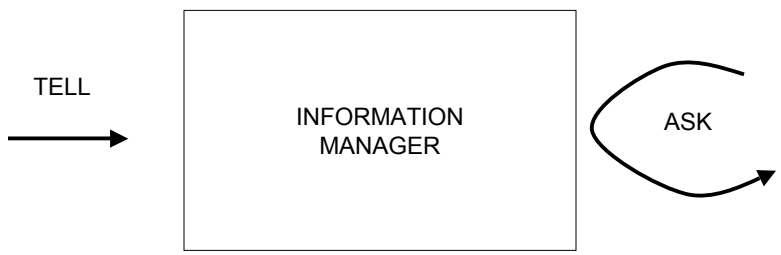
Recap

Week 1: Motivation for info management
- Value of information about the world

Week 2: Theory of representation
- Symbols can correspond to entities and relationships
- Do logic on symbols

Week 3: Practice of representation
- What really happens now?

Pictorial view



Example, more concrete

Step 1: **IM =** ∅

Step 2: **TELL(IM, f(a,c))**

- Give IM the information that the individual represented by **a** stands in the relation represented by **f** to the individual represented by **c.**
- **Hey, Toyota Prius #NJ YY-901 just went through the exit 9 toll plaza!**

Example, more concrete

Step 1: **IM =** ∅

Step 2: **TELL(IM, f(a,c))**

Step 3: **ASK(IM, f(X,c))**

- Ask IM to report proofs that show that some individual **X** stands in the relation represented by **f** to the individual represented by **c.**
- **Hey, what cars went through the exit 9 toll plaza just now?**

Example, more concrete

Step 1: **IM = ∅**

Step 2: **TELL(IM, f(a,c))**

Step 3: **ASK(IM, f(X,c))**

- Ask IM to report proofs that show that some individual **X** stands in the relation represented by **f** to the individual represented by **c.**
- IM answers with a single proof of the form

  f(a,c) → f(X,c) [X=a]
- **It was Toyota Prius #NJ YY-901!**

Modeling

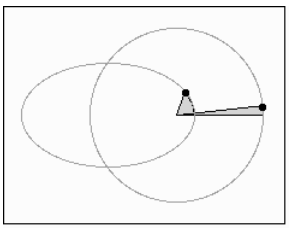
Practice of creating a model

- Description of the real world
- Specified precisely
- Only approximately correct

Mathematical Modeling

Like Kepler's laws of motion

$$t^2 = c\, d^3 \quad \text{(third law)}$$



(GIF from D. Koch, NASA, curator Kepler project)

Kepler's Model

Precise description of the real world

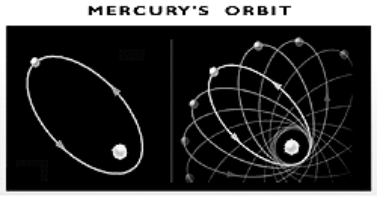
- Could enable explicit calculation of positions
- Accurately predicted positions of planets over decades

Only approximately correct

- Actually, small deviations in orbit of Mercury

Approximation

Actually need general relativity



Mercury's orbit precesses

- Perturbation from planets: 541 arcsec/100yr
- GR perturbation: 43 arcsec/100yr

See phys FAQ http://math.ucr.edu/home/baez/physics/index.html

Image from Einstein's legacy, NCSA

Conceptual modeling

Modeling based on concepts not equations

- Qualitative, binary distinctions among kinds of things
- Inspired or informed by commonsense ways of thinking about or ways of talking about the world

Yahoo Directory

Conceptual model of the web

– Organized by topic, e.g., top-level categories

Business	Regional
Computers	Society
News	Education
Entertainment	Arts
Sports	Science
Health	Social science
Government	Reference

Examples

Includes taxonomic information

– Literally, classification of living organisms:

Science > Biology > Animals > Arthropods > Crustaceans > Lobsters

– More generally, locating a category uniquely within a hierarchy:

Regional > States > New Jersey > Cities > New Brunswick

Examples

Also includes crosscutting views

Science > History > Mathematics > Pi

is a “link” to

Science > Mathematics > Numerical Analysis > Numbers > Specific Numbers > Pi > History

Yahoo methodology

Yahoo looks at web sites

- People submit their own sites for indexing (For-profit sites cost $300)
- Apparently, some issues:

  We [Haystack in a Needle Consultants] highly recommend using a professional service to handle your Yahoo submission. Changing a directory listing on Yahoo is next to impossible, so you want to make sure your listing has been optimized for keyword density before submitting.

Yahoo use

Indexing sites by categories

- Yahoo delivers 15% of search traffic

Conceptual modeling in CS

Artificial intelligence – cognitive science

- Describing human knowledge
- Simulating human mental processes
- Engineering intelligent behavior

Software systems

- Requirements analysis
- Database design
- Decision support

Today

Conceptual modeling

- What concepts are there?
- How are they related?
- How do you find them?
- How do you represent them for DBs?

What concepts are there?

Concepts are specific

What does it mean to be red

- Red pen
- Red shoes
- Red apple
- Red grapefruit
- Red traffic light
- Red wine
- Red state

Concepts are specific

What does it mean to be tall

- Tall mountain
- Tall building
- Tall ship
- Tall basketball player
- Tall table
- Tall latte
- Tall tale

Specificity

All these properties have something in common

- Underlying generalization
- Determines aspects of appearance
- Along a certain dimension

But for a conceptual model, need to specify one case.

Why?

Semantics

- Want clear criteria about whether statements using symbols are true or false

Inference

- Want rules like this to work…

  If X is tall and Y is short then
  the height of X is bigger than the height of Y

How do you deal with specificity?

Words depend on context

Need guidelines that spell out the context

- General principles
- Common mistakes
- Good examples

What Yahoo has to settle

How many people do you have to kill before you are a serial killer

Answer:

- Three.
- Also, it can't be a spree.

http://ask.yahoo.com/ask/20050204.html

Better Example: UMLS

Unified medical language system
- Started in 1986
- NLM long term R&D project
- For retrieving medical information

Current status:
- 1 million distinct concepts
- 5 million terms from specialized sources
- 9 million relationships among concepts

UMLS Example
(Separate slides from Tilley & Willis)

Further information:
- http://www.nlm.nih.gov/research/umls/
- http://www.nlm.nih.gov/research/umls/umlshelp.html
- http://umlsinfo.nlm.nih.gov/styfile.html

UMLS: Social Behavior

Definition:

- Behavior that is a direct result or function of the interaction of humans or animals with their fellows.

Note in definition:

- This includes behavior that may be anti-social.

UMLS: Social behavior

Examples:

- Acculturation; Communication; Infanticide; Interpersonal Relations; Social Conformity

UMLS: Social behavior

Usage note:

- 'Social Behavior' requires the direct participation of others and is, thus, distinguished from 'Individual Behavior' which is carried out by an individual, though others may be present.

How are concepts related?

How are concepts related?

isA link: X isA Y

- Every instance of category X is an instance of category Y
- Genus and differentiae
  Taxonomic subclasses have no overlap

Ancient Semantic Net: *Tree of Porphyry*

3rd C. AD

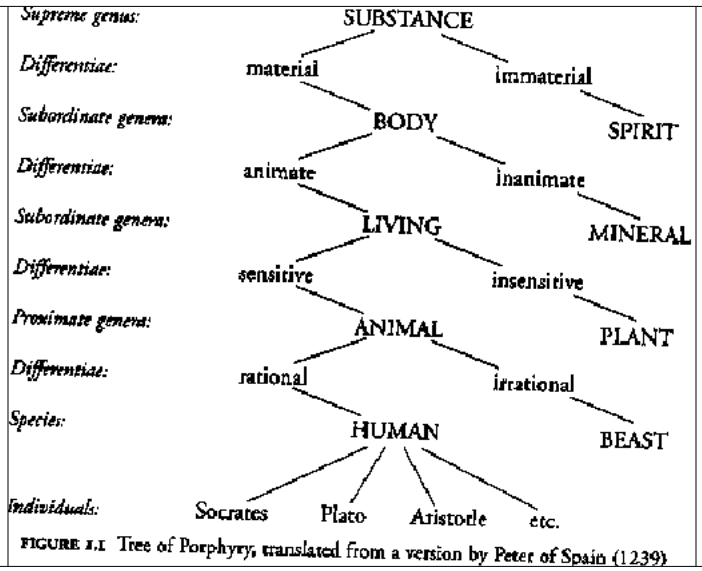


FIGURE 1.1 Tree of Porphyry, translated from a version by Peter of Spain (1239)

(from J.Sowa,"Knowledge Representation")

A network of human concepts?

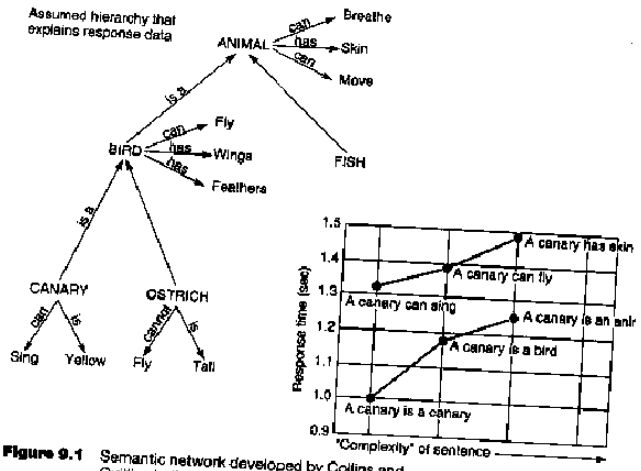
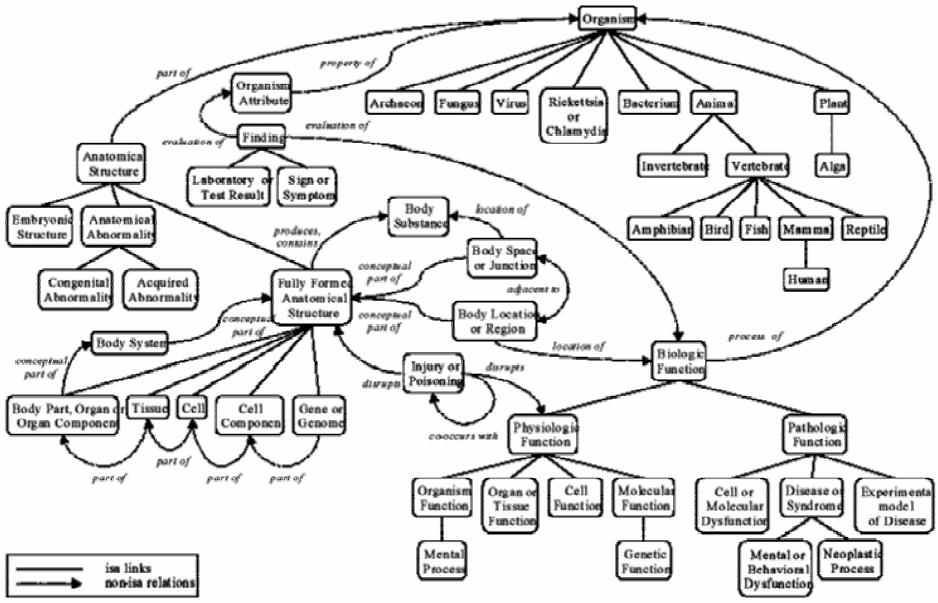


Figure 9.1 Semantic network developed by Collins and Quillian in their research on human information storage and response times (Harmon 1985).

Other relations

54 relations in UMLS, such as:

- **Therapeutic or Preventive Procedure | isa | Health Care Activity**
- **Therapeutic or Preventive Procedure | treats | Injury or Poisoning**
- **Behavior | result_of | Mental Process**
- **Mental Process | process_of | Vertebrate**
- **Virus | causes | Pathologic Function**



Relations are also specific

Example: Parts and wholes

- Easily confused with isA – hierarchical relationship is usually obvious in context

Computer — isKindOf
Microcomputer — isPartOf
Mother_board — isPartOf
CPU — isPartOf
Pentium — isKindOf
Transistor — isPartOf

Kinds of part-whole relationships

Complex:component
- Box – lid

Collection:members
- Shipment – box

Mass:quantities
- Ocean – bucket of seawater

Closely related relationships

isContainedIn
- Gift – box

isConnectedTo
- Carriage – horse

isBranchOf
- Iliac artery – aorta

hasLocation
- Rutgers – Piscataway

Zeroing in on the intended relationship

Transitivity

- The finger nail is part of the finger is part of the hand is part of the upper extremity is part of the body.

Contrast:

- The foot of the goose is part of the goose but not part of the flock of geese.

Zeroing in on the intended relationship

A fault to the part is a fault in the whole

- Injury to the fingernail is injury to the body
- A fault in the tail light is a fault in the car

Contrast:

- A fault (e.g. souring) to the milk contained in the bottle is not damage to the bottle

Zeroing in on the intended relationship

Existence:

- part may depend on the whole for its existence (e.g., chapter of a book)
- whole may depend for its continued existence on a part that is irreplaceable (e.g., book on its title -- if you changed the title it would be a different book)

Even more specificity: Kinds of parts

isComponentOf

- "the leg is a component of of the table"
- Discrete, connected, clear boundary, specifically named, may be differently constituted (can have metal legs on a wooden table or vice versa)

Even more specificity: Kinds of parts

isSubdivisionOf

- 'France is subdivision of Europe"
- Arbitrary, similarly constituted, components typically fall into one or another subdivision; defined in relation to something else; sensible to talk about what fraction it is: half the table, a third of the table, etc.
- Interesting 'transitivity' note: Components of subdivisions are components of the whole, but subdivisions of components are not subdivisions of the whole

Even more specificity: Kinds of parts

isFunctionalPartOf

- 'The remote control is part of the projection system'
- Part of a common function; while structural parts form a contiguous whole, they may or may not contribute to function e.g. decorative parts

How do you find them?

See what's out there!

- Crawl the web
- Search using a web index
- Ask domain experts – like doctors
- Collect suitable data yourself

WordNet

http://www.cogsci.princeton.edu/cgi-bin/webwn

Thesaurus of word meanings "senses" plus ability to find

- synonyms
- coordinate terms (siblings in IsA hierarchy)
- hypernyms (IsA ancestors)
- hyponyms (IsA descendants)
- holonyms (isPartOf)
- meronyms (hasPart)

Computational Lexicography

A 'word sketch' is a summary of the interesting collocations and grammatical patterns a word occurs in, produced automatically from a large corpus, designed to help a lexicographer produce an accurate dictionary entry for the word.

http://www.itri.brighton.ac.uk/~Adam.Kilgarriff/wordsketches.html

BNC

The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of current British English, both spoken and written.

http://www.natcorp.ox.ac.uk/

How do you represent them?

Big Picture

Issues in ER design

Modeling individuals

Many kinds:

- **_concrete_: Janet, that tree**
- **_abstract_: number 12, Rutgers, cs336**
- **_hypothetical_: Santa, King of USA**

Modeling individuals

Granularity

- Book vs book edition vs book copy
- Flight 303 from Newark to Toronto
- Flight 303 on February 3, 2004

Modeling relationships

Individuals & relationships:

- Dana hasBorrowed book23

Binary relationships have inverse relationship; often it has a name:

- book23 lentTo Dana

Tip: name relations asymmetrically

- (e.g., not "loan"),
- R(a, b) should be read as "a R b" (I tend to use verbs, or has___, or __Of as names)

Modeling Categories

An important motivation for categories is to constrain relationships

- *domain*: **only** Material **can be** lentTo

- *range*

may be relative to domain:

Material *can be* lentTo Patron

Journal *can be* lentTo Faculty

Modeling subcategories

Disjoint subclasses frequently factor a class

– Partition of material by medium:

Print, Audio, Video => Material

– Partition of material by format:

Book, Journal,...=> Material

Some subclasses cover the super-category

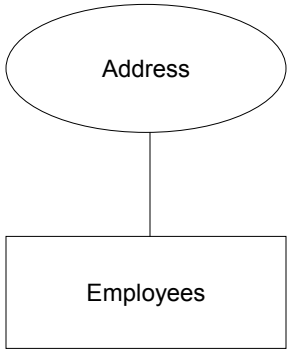
– every instance of C is in one of C's subclasses

Modeling subcategories

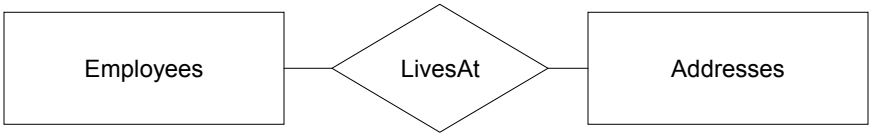
Tip: Construct a “backbone” hierarchy of primitive, disjoint concepts, with each split providing concepts at the “same level”

- Exploit inheritance
- Be on the lookout for specialized restrictions
- e.g., though Person might have restriction **(age between 0 and 120)** Employee might have restriction **(age between 16 and 65)** reflecting mandatory retirement and laws against child labor.

Entity versus Attribute



Entity versus Attribute

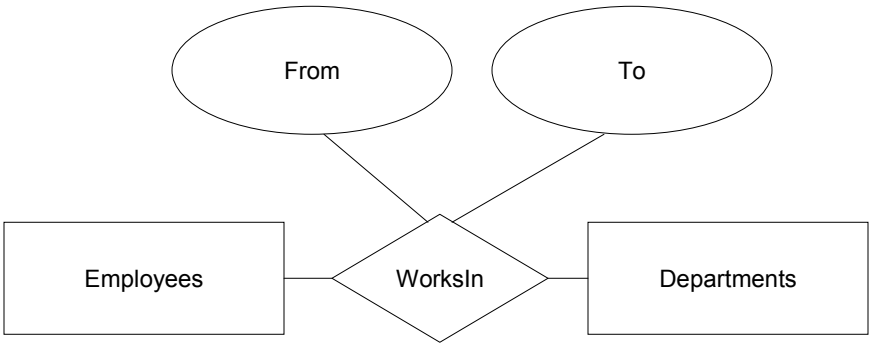


Entity vs Attribute

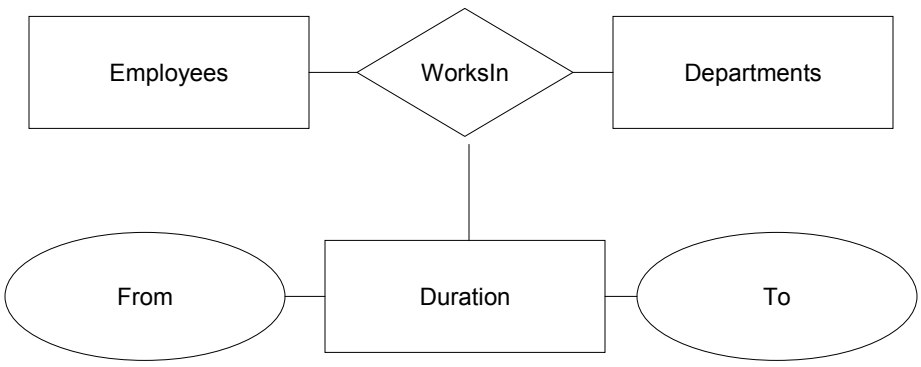
Entity

- Can have structured representation
  like fields of address
- Allows multiple relationships
  more than one address per employee
- More complicated representation

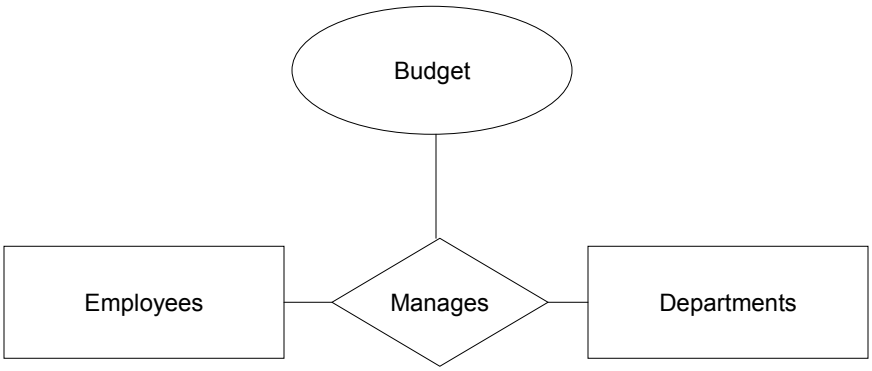
Entity vs Attribute



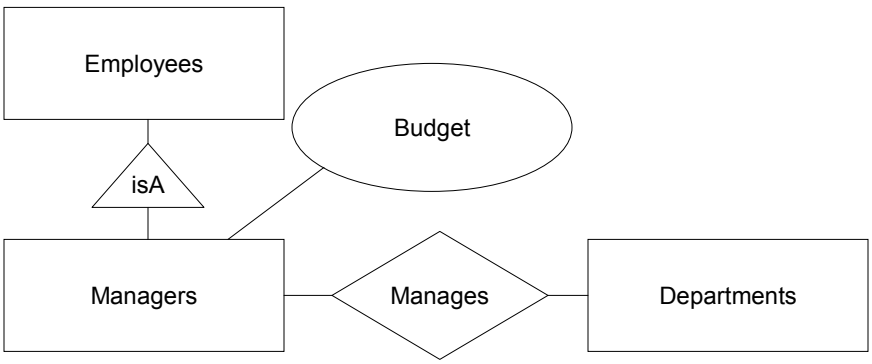
Entity vs Attribute



Entity vs Relationship



Entity vs Relationship



Entity vs Relationship

Attribute of relationship

- Associates attribute with relatoinship
- If attibute really belongs with entity in the relationship, need to create a new entity set.