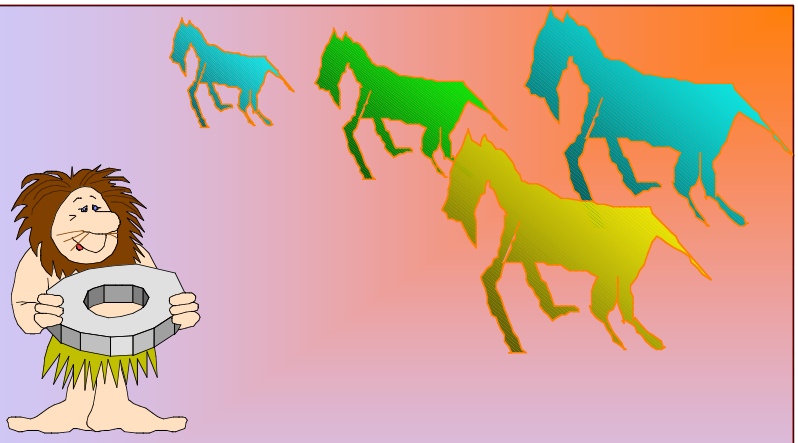
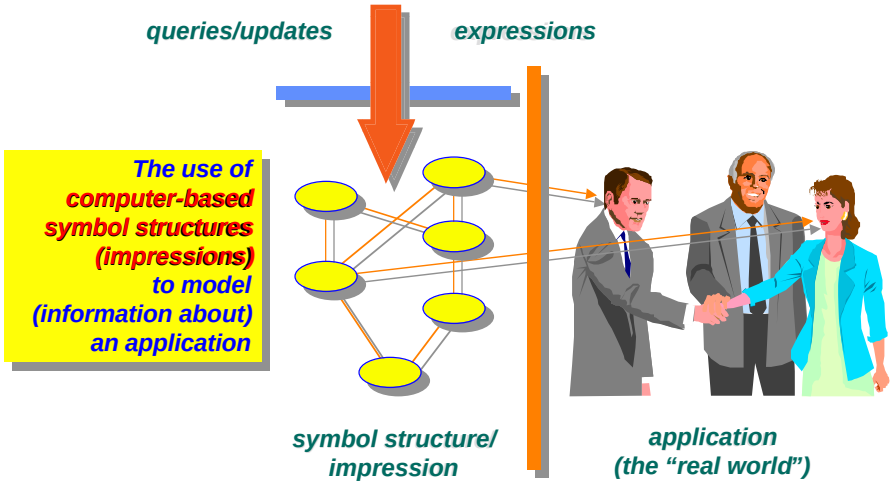
I. Introduction

What is Information Modeling?
Impressions and Expressions
Conceptual Models
Primitive Terms, Abstraction Mechanisms and Tools
Examples of Conceptual Models:
EER, UML, CLASSIC, KAOS, Telos, …and Tropos
History of Conceptual Modelling
Formal and Informal Models



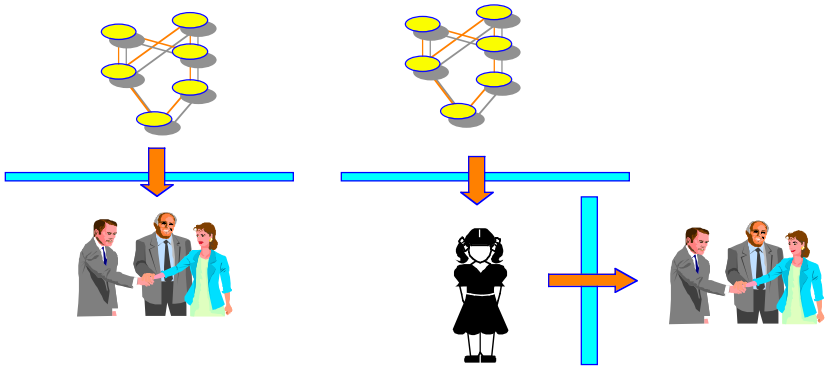
Humans used symbols to model their environment since the beginning of civilization!

What is Information Modeling?



...Actually...

We could be modeling the real world, or **somebody's conception** of the real world



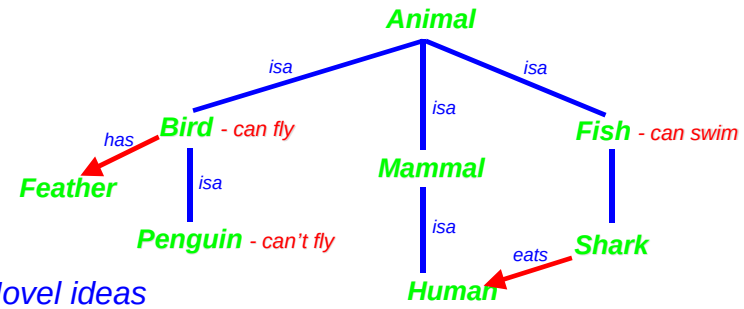
Information Modeling

- The term "information modeling" is actually ambiguous.
- Consider
  - ✓ "Designer clothing" -- clothing made by designers (subjective reading); hence "information modeling" means "models made out of information";
  - ✓ "Traveller clothing" -- clothing for the traveller (objective reading): hence "information modeling" means "models of information";
- In general, we will use the second reading: we are interested in models of information (about the world.) The information may come from a single agent, or a group of agents.

Origins

- Ross Quillian proposed ***semantic networks*** as a model of the structure of human memory (1966)
- Ole-Johan Dahl proposed in 1967 ***Simula***, an extension of the programming language ALGOL 60, for simulation applications which require some "world modeling"
- Jean-Robert Abrial proposed a ***semantic model*** in 1974, shortly followed by Peter Chen's ***Entity-Relationship model*** (1975) as advances over logical data models, such as Codd's Relational model.
- Doug Ross proposed in the mid-70s the ***Structured Analysis and Design Technique*** (***SADT***) as a "language for communicating ideas". The technique was used to specify requirements for software systems.

Semantic Networks (1966)



Novel ideas

- Impressions are built out of ***concepts*** and ***associations***
- ***Inheritance of attributes*** -- default, single
- Computation defined in terms of ***spreading activation*** -- e.g., discovering the meaning of "horse food"

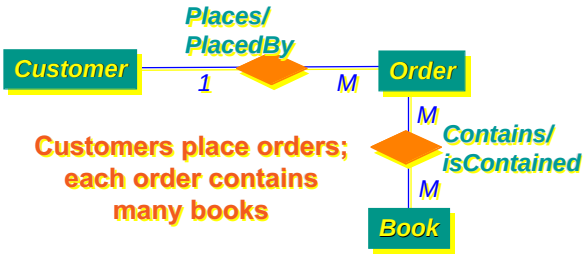
  horse --> animal --> eat --> food

  horse --> animal --> madeOf --> meat --> food

Simula (1967)

- Ole-Johan Dahl proposed it as an extension of the ALGOL 60, for simulation applications.
- A (simulation) program consists of classes and instances.
- Instances ***model the simulated application***, classes define common features of instances, are organized into subclass hierarchies.

The Entity-Relationship Model (1975)



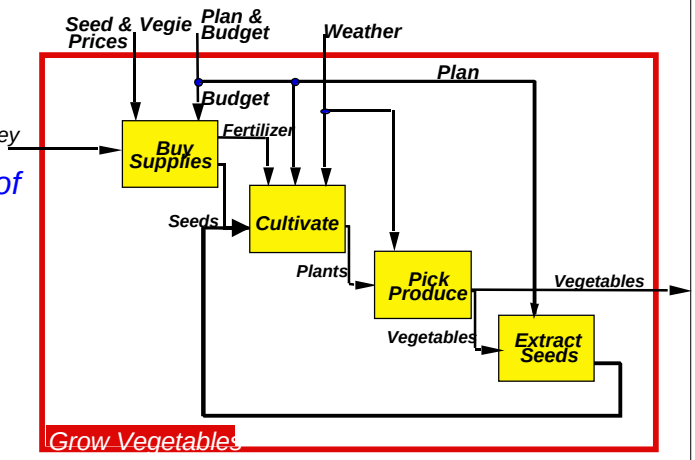
Novel ideas

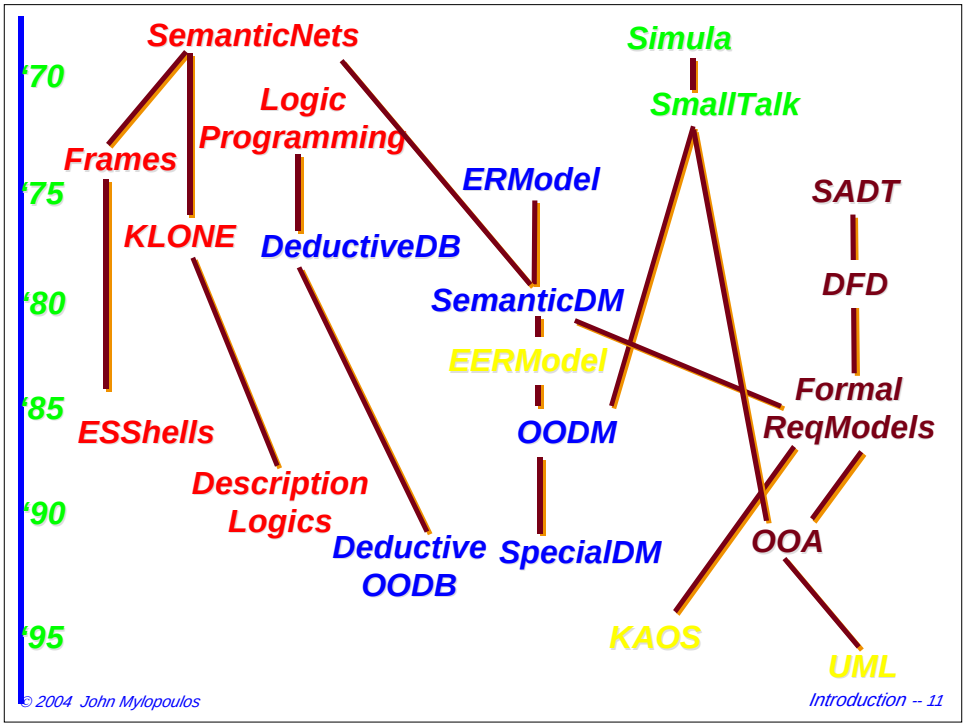
- *Assumes that application consists of **entities** and **relationships** (**ontological assumptions**)*
- *Shows how a conceptual schema can be mapped onto a logical one.*
- *[Abrial's semantic model was more akin to OO data models, but did offer entities and relations too]*

Structured Analysis and Design Technique (SADT)

Novel Ideas

- *Modeling the operating environment of a software system.*
- *Application models organized in terms of box-inside-box notation.*





Different areas of Computer Science use different terms for the same concepts!

Can we find a common vocabulary? (...please??)

Expressions

- ***Expressions*** *are communication statements of linguistic and/or pictorial nature through which an agent (human or otherwise) accumulates information about an application.*
  - ✓ *Natural language statements are expressions.*
  - ✓ *So are books...*
  - ✓ *So are Predicate Calculus statements...*
  - ✓ *So are SQL statements....*
  - ✓ *So are digitized pictures...*
  - ✓ *So are Web pages...*
  - ✓ *...more...*

Impressions

- ***Impressions*** *(also* ***models****) are internal symbol structures held by an agent (human/software) that represent fragments of an application.*
- *Example impressions: human memory, mental states, databases, knowledge bases, websites,...*
- *Impressions have two properties:*
  - ✓ ***Functional role*** *-- impressions arise by virtue of the history and coupling of the agent to its environment, and give rise to the agent's future actions;*
  - ✓ ***Representational import*** *-- impressions stand in a content relationship to an application the agent is in contact with.*

Meaning vs Interpretation

- The ***meaning*** of an impression is what all uses of the impression have in common. For example, "the morning star" always refers to a star visible early in the morning.
- The ***interpretation*** of an impression is what the impression refers to in a particular instance, or a particular use. For example, "the morning star" may refer to planet Venus when I use it (...on Earth), while it refers to planet Jupiter when astronaut John Glenn says it on Mars.
- Typically, the transition from meaning to interpretation occurs when an impression is bound to a context. This applies to indexical terms, e.g., "I", "here", "now", but also descriptions, such as "morning star" and "a student in CSC2507".J

Information Modeling in Computer Science and Engineering

Information modeling is practiced and researched in several areas within Computer Science and Engineering, including:

- **Databases** -- data modeling, using vanilla or semantic data models to build ***databases***;
- **Software Engineering** -- requirements modeling, using diagrammatic (structured, OO, or plain vanilla) techniques to build ***requirement specifications***; software process modeling using finite state machines, statecharts, rules, Petri nets to build ***process models***;
- **Artificial Intelligence** -- knowledge representation using logic-based notations, description logics, semantic networks, frames, etc. to build ***knowledge bases***;

Variations in Use

- There are important differences in the use of impressions among different areas of Computer Science:
- In AI, knowledge bases are to be used by ***a program*** to perform a task (expertly).
- In Software Engineering, diagrams are used for communication ***among people***.
- Databases model large amounts of concrete facts, rather than generic knowledge, are used by ***humans and programs*** and have to scale up.

Implications

Differing uses of impressions have important consequences

- ***Type of notation used*** -- if people use the impression, the notation may be semi-formal or even informal; for knowledge bases in AI, on the other hand, formality is all-important.
- ***Types of knowledge captured*** -- for SE and Databases domain information is captured; for AI, task knowledge, including heuristics is equally important.
- ***Coverage and completeness of the impressions*** -- for SE and Databases coverage can be quite broad and incomplete; for AI, coverage has to be narrow and complete.

Information Models

- An ***information model*** consists of:
  - a collection of symbol structure ***types***, whose instances are impressions that describe an application, a collection of ***operations***, and a collection of general ***integrity rules*** which define the set of consistent impressions, or changes thereof**.**
- Example: The Relational Model for databases
  - Basic impression types: table, tuple, domain
  - Operations: add/remove/update tuple,
    - union/ intersection/...of tables
  - Integrity rule: No two tuples within a table can have the same key

Types of Information Models

- ***Physical models*** use machine-oriented terms -- e.g., records, fields, B-trees,...-- to model an application.

  ...conflicting representational and efficiency concerns!
- ***Logical models*** offer mathematical abstractions -- e.g., arrays, lists, sets, relations -- for modeling purposes, hiding the implementation details from the user.

  ...but logical symbol structures are flat and unintuitive....
  can't describe complex situations!
- ***Conceptual models*** use application-oriented terms and organize information on the basis of principles such as generalization, aggregation and classification.

  conceptual symbol structures (are supposed to)
  model directly and naturally an application

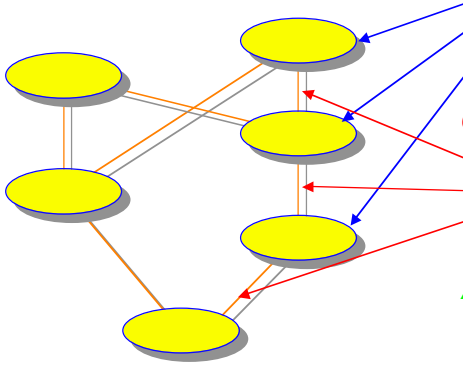
Information Models, Languages and Notations

- Unlike an information model, a ***language*** is used for communication.
- There may be several languages associated with one and the same information model.

  e.g., relational query languages
- A ***notation*** (usually graphical) is used to offer partial descriptions of an impression. Again, there may be several notations associated with the same information model.

  e.g., graphical notations for Data Flow diagrams.

Characterizing Conceptual Models



Basic building blocks -- Terms
descriptions
concepts,...

Composition rules -- Abstractions
abstraction mechanisms
semantic relationships
structuring mechanisms,...

Analysis and management tools
simulation,
verification,
normalization,...

Terms, Abstraction, and Tools

- Let's try to characterize conceptual models by looking at the basic building blocks, the structuring mechanisms, and the tools they offer for building an information base:
- **Primitive Terms** - these are the concepts built into a conceptual model, e.g., **Entity**, **Activity**, **Agent**, **Goal**,...
- **Abstraction Mechanisms** -- primitive mechanisms for structuring an impression along different dimensions, e.g., **Generalization**, **Aggregation**, **Classification**,...
- **Tools and Analysis techniques** -- for creating, updating, searching, validating and managing an information base

Important to avoid superficial comparisons,

e.g., ones based on syntactic/graphical sugar

Three Conceptual Models

- **Extended Entity-Relationship Model** (**EER**) -- ER model extended to support aggregation and generalization; there are many different version of this, e.g., [EER92]; some versions are supported by commercial modeling tools.
- **Unified Modeling Language** (**UML**) -- combines earlier Object-Oriented Analysis (OOA) techniques by Booch, Jacobson (OOSE), Rumbaugh (OMT), others [UML97]; offers facilities for modeling objects, methods, various types of actors, aggregation, generalization, etc.
- **CLASSIC** -- A Description Logic developed at Bell labs by Brachman, Borgida etc. [Borgida89]; uses a limited form of Logic, supports a limited form of inference.

...And Three More...

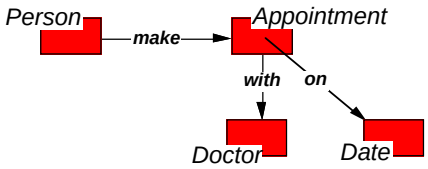
- **KAOS** -- *a research prototype for requirements modeling, developed by Axel van Lamsweerde and colleagues [Dardenne93]. Can model goals, constraints, tokens, classes, metaclasses etc.*
- **Telos** -- *a research prototype intended for metamodelling applications [Mylopoulos90]; treats attributes as first class objects, uses heavily metaclasses (for objects and attributes) for metamodelling, supports a logic-based sublanguage for specifying constraints and deductive rules.*
- **Tropos** -- *a modeling language founded on the notions of actor, goal, and social dependency (among actors); used for modeling different phases of software development.*

Formality: From Informal to Formal

Informal models, minimal syntax, no primitive terms or semantics	e.g., graphs
Informal models, minimal syntax, primitive terms, no semantics	e.g., SADT, DFDs
Formal models with primitive terms, syntax and semantics	e.g., EER, some OOA models
Formal models, with a primitive terms, syntax, semantics and an assertional sub-language for user-defined constraints and rules.	e.g., RML, KAOS

Informal Models

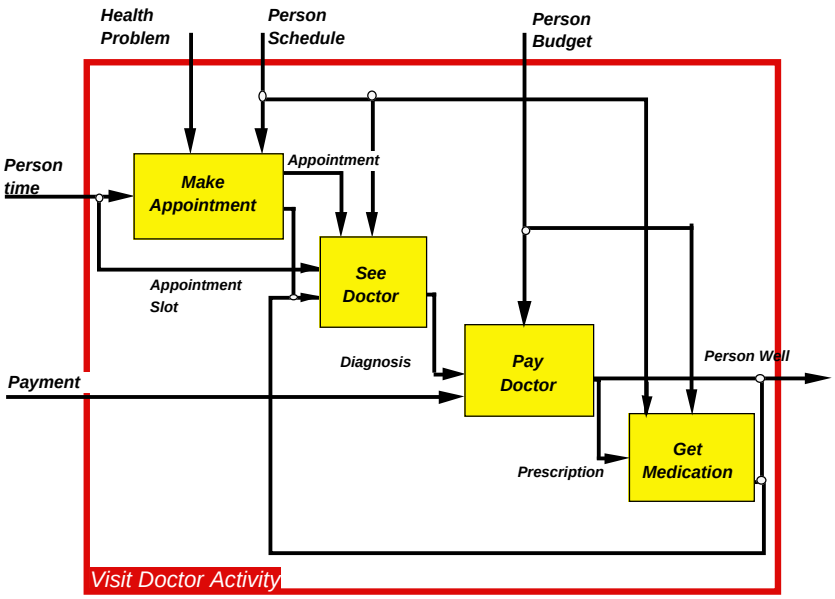
- Diagrammatic notations for sketching an application.
- Offer boxes (nodes) and labelled arrows (links).
- There is no guidance on what boxes and arrows to use for a particular application.
- There is no agreed upon semantics attached to either the boxes or the arrows; users interpret them in terms of their labels, other informal documentation, as well as their own perspective



Diagrammatic Models

- Offer a graphical notations for modeling; they come with primitive terms for modeling an application, i.e., a set of concepts in terms of which we can conceptualize an application

  e.g., SADT offers **data** and **activity** boxes, **input**, **output**, **control** arrows
- There are only **informal semantics** attached to either the boxes or the arrows; users interpret them in terms of their labels, other informal documentation.
- Such techniques usually come with a methodology on how to use them.



Formal Models

- Used as far back as semantic networks [Findler78], also in semantic data models, such as the Entity-Relationship model.
- We could provide a formal semantics for the SADT box and arrow types in terms of a set of axioms (or in some other way)

  $\forall$ A/Activity, E/Entity, t, t'/TimeInterval

  [InputOf(E, A) $\wedge$ InstanceOf(a, A, t) $\Rightarrow$

  $\exists$ e InstanceOf(e, E, t') $\wedge$ Overlaps(t', t) $\wedge$ InputOf(e, a)]

  where Overlaps(t', t) means 

- Such an axiom might be enforced by an SADT++ tool which instantiates activities and data

Analysis of Impressions

- *Formality is important for an information model for two basic reasons:*
  - ✓ *It eliminates, or at least reduces, ambiguity;*
  - ✓ *It can be used as a foundation for (algorithmic) analysis of impressions.*
- *Analysis may consist of various forms of consistency, completeness, etc.*
- *The course will be discussing different forms of analysis for different kinds of impressions.*

References

- [Borgida89] Borgida, A., Brachman, R., McGuiness, D., and Resnick, L., "CLASSIC/DB: A Structural Data Model for Objects", Proceedings ACM SIGMOD International Conference on the Management of Data, Portland, June 1989.
- [Dardenne93] Dardenne, A., van Lamsweerde, A. and Fickas, S.,"Goal-Directed Requirements Acquisition", in *The Science of Computer Programming 20*, 1993.
- [EER92] Parent, C., and Spaccapietra, S., "ERC+: An Object-Based Entity-Relationship Approach", in [Loucopoulos92].
- [Loucopoulos92] Loucopoulos, P. and Zicari, R., (eds.) *Conceptual Modeling, Databases and CASE: An Integrated View of Information System Development*, Wiley, 1992.
- [Mylopoulos92] Mylopoulos, J., "Conceptual Modeling and Telos", in [Loucopoulos92].
- [Smith86] Smith, B., "The Correspondence Continuum", Proceedings Canadian Society for the Computational Studies of Intelligence (CSCSI) National Conference, Montreal, 1986.