On difficulties of forming opinions on what you don't know that you don't know - in Information Systems Engineering

Arne Sølvberg
Dept. Information & Computer Science
NTNU-The Norwegian University of Science and Technology
Trondheim
Norway

Information Systems must be built to support continuous change

- IS to support non-routine problem solving
- Routine processing can be pre-defined
- Non-routine problem solving organizations learn as they operate
- Non-routine problems change as they are solved; we get into unchartered territory
- IS support must adapt to the changes
- ISE must support emerging workflows

What is Engineering?

- Technique is physical means to satisfy human desire
- Technology is the discipline of the techniques, their means and their use
- Artifacts are systems of techniques (= systems of physical means that satisfy human desires), e.g., clocks, automobiles
- Engineering is the process that creates artifacts

What is Information Systems Engineering?

- Information Systems Engineering is the application of computers (the physical means) in order to satisfy human desires
- ISE is about defining goals for human desires
- ISE is about finding computer based solutions that satisfy the agreed goals
- ISE is about defining a solution space that satisfies the agreed goals
- ISE is about finding a “point” in the available solution space that gives maximal benefit/cost

Are human desires independent of the available means to satisfy them?

If not,

then

- goals depend on the means, and
- goals depend on the human understanding of how to apply the means

and

- goals have to be continuously evaluated and reformulated as we learn more about
  - the consequences of reaching the goals, and about
  - other reachable and desirable goals

A view of social systems, according to George Soros

- Social situations have thinking participants who influence the situations in which they participate
- Thinking participants understanding of the world consists of knowledge and beliefs (= a mental model of the world)
- Knowledge consists of facts, which are true
- Understanding may approximate truth, but can never qualify as knowledge
- The divergence between reality and our understanding of the world is itself part of reality
- Reality is a moving target that remains just beyond our range forever

Reflexivity – the cornerstone of George Soros' conceptual framework

- *Participation* and *understanding* interferes with each other, ensuring that
  - Our understanding is inherently imperfect, and
  - Our actions have unintended consequences
- **Reflexivity** is the two-way connection between thinking and reality

“Radical Fallibility”

- Popper: We may be wrong. All scientific statements must be subjected to falsification by testing
- Soros: We are bound to be wrong in some way or another (the postulate of radical fallibility)
- “fertile fallacies”: an idea valid in one field is extended to fields where it no longer applies, where it stimulates to the development of new ideas
- Soros: “Our whole civilization is a product of fertile fallacies”

“Open society”

- As participants in social situations we base or actions on our understanding of society
- BUT : we may be wrong
- So, we must treat our understanding as being provisionally true, and keep it open to continuous re-examination
- This is the foundation principle of an Open Society, and keeps it open to improvement

“The Human Uncertainty Principle”

- Fallibility affects not only our thinking but also the reality we try to understand
- Fertile fallacies can go along way in shaping the world in which we live
- E.g., a theory of physics does not change the material world, but a theory of society may change society
- The radical fallibility of participants introduces uncertainty in the situations in which they participate = the human uncertainty principle
- If our understanding is inherently flawed, the extent of our misunderstanding becomes all-important

Communication is the key!

- For people to communicate without distortion requires shared understanding of reality
- Shared understanding requires shared mental models of the world
- Mental models are conceptual models
- Individual mental models develop through communication with others
- A shared language for conceptual modeling?

Data, Information, Knowledge: The infological equation

Let m be a message from a sender to a receiver
Let t be the time of sending the message
Let S be a mental model of a system
Let i be an interpretation function
Let I be the information content of the message m
The infological equation states that

I=i(S,m,t)

The meaning of a message received depends on the mental domain model of the receiver.
The intended meaning depends on the mental domain model of the sender.

Information is the interpretation of the data of a message together with the domain knowledge of the receiver/sender

Basic Concepts

- *Data* are digital numbers, and are interpreted both as numerals and linguistic units like words and phrases.
- *Knowledge* consists of *statements* that correspond to *facts* which are considered to be true.
- *Beliefs* consist of statements that may not be true, in the sense that they correspond to facts.
- An individual's *understanding* (of *reality*) consists of that individual's knowledge and beliefs.
- Knowledge and beliefs of other human beings are parts of *reality*.
- *Messages* consist of *data*.
- *Meaning of a message* is a decoding of the message into statements of knowledge and belief.
- *Information* (for an individual) is an *interpretation of a message received*, where the interpretation has the capacity of changing that individual's understanding (mental model) of the world.

Information systems are too complex to be understood in every detail

Information systems are

- Complex, unsurveyable, imperceivable
- Not to be understood in every detail by single individuals

Donald Rumsfeld, while Defense Secretary to Bush the Son:

- There are things that you know
- There are things that you know that you don’t know
- There are things that you don’t know that you don’t know

There are things that you know that you don't know, an example

Let **most, tall** and **blond** be fuzzy variables.

Information base:

Most Swedes are tall.
Most tall Swedes are blond.
Magnus is a Swede.

Question:

What is the probability that Magnus is blond?

Answer: at least **most x most**

What about what we don’t know that we don’t know in ISE?

- Information Systems Engineering offers an approach to uncovering the unknown
- The approach is common to all engineering
- The approach is one of systematic iteration between
  - solution proposal evaluation and
  - requirement modification (goals, restrictions)

Important issues in ISE

- information system project management
- systems design approaches
- modeling languages for information systems, domain systems and data systems
- comprehensibility of specifications
- modeling of the meaning of data
- validation techniques
- coping with changes in the domain and in the technological environment

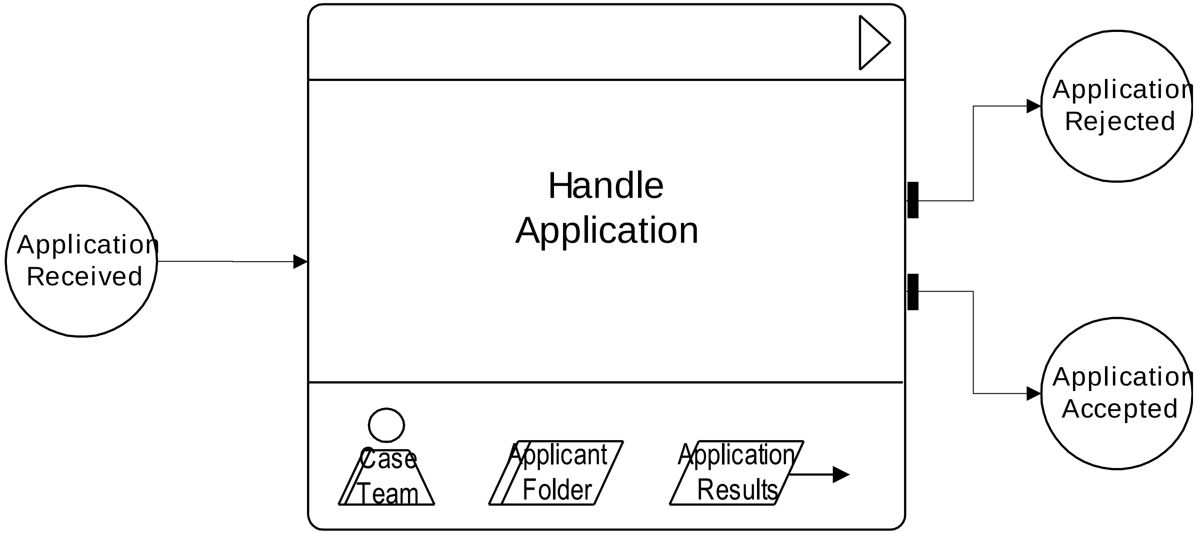
Requirements to information systems engineering

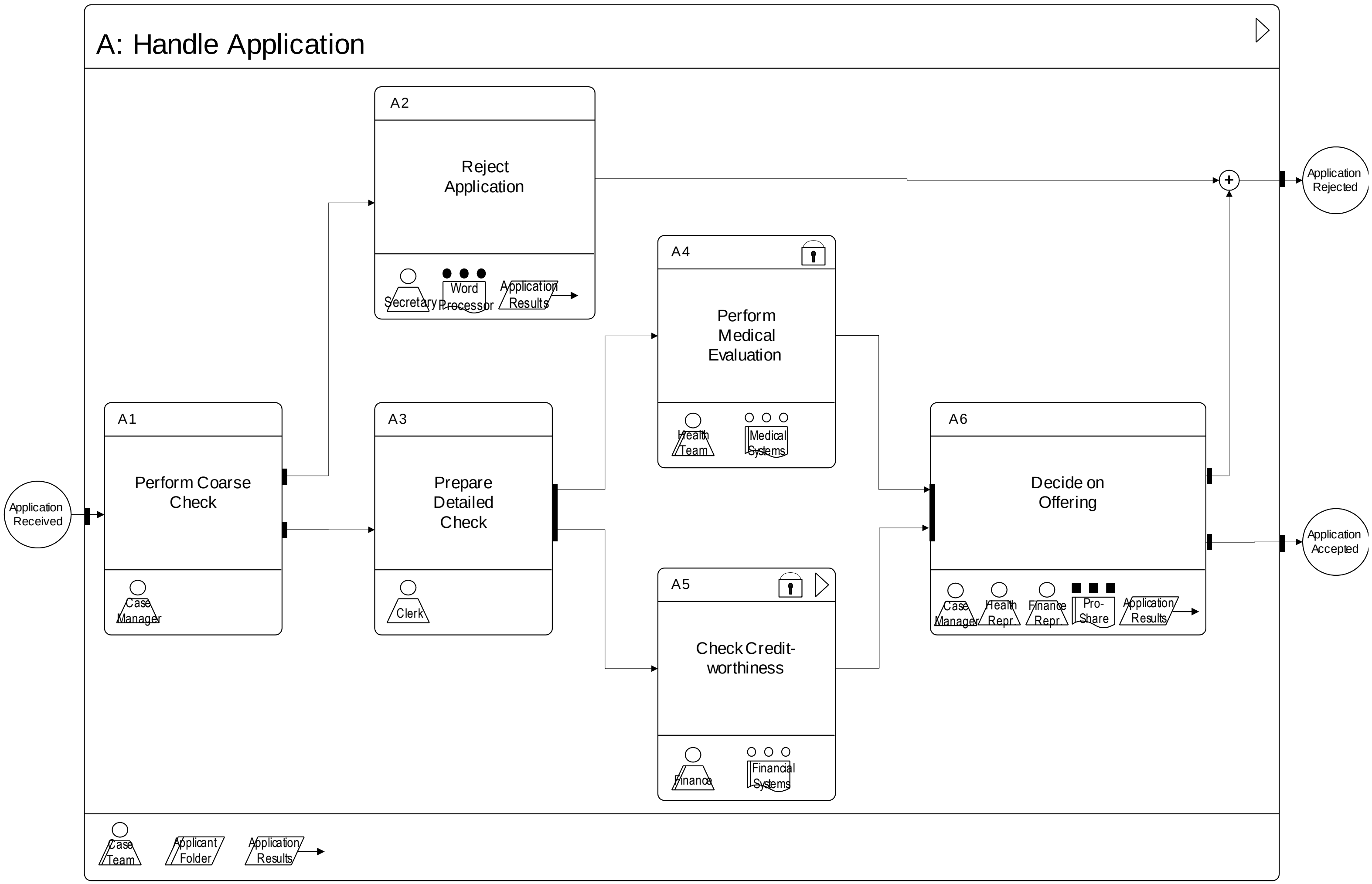
- the theoretical framework must make it possible to build *integrated models* of systems which consist of both digital and non-digital components
- information models must represent the *meaning* of data, that is, data should be explicitly related to the phenomena they represent
- system models must be *comprehensible* on every level of systems detail
- system models must permit *specification in terms of solutions* on every level of detail, in order to provide for executable specifications
- system models must support the need for *validation* of design proposals during their development
- system models must support *different specification detail*, both formal and informal specifications
- system models must encourage *systematic evolution of specification detail* from low level of detail to high level of detail
- system models should support proven *engineering practice*

The APM modeling language

- IS-support for non-standard problem solving
- “Emerging” workflows
  - IS support must reflect changes in problem
  - Resources must be reallocated
  - Knowledge must be re-used
- Is there a solution?

Workflow essentials





A: Handle Application
A1
Perform Coarse Check
Case Manager
A2
Reject Application
Secretary
Word Processor
Application Results
A3
Prepare Detailed Check
Clerk
A4
Perform Medical Evaluation
Health Team
Medical Systems
A5
Check Credit-worthiness
Finance
Financial Systems
A6
Decide on Offering
Case Manager
Health Repr
Finance Repr
Pro-Share
Application Results
Application Received
Application Rejected
Application Accepted
Case Team
Applicant Folder
Application Results

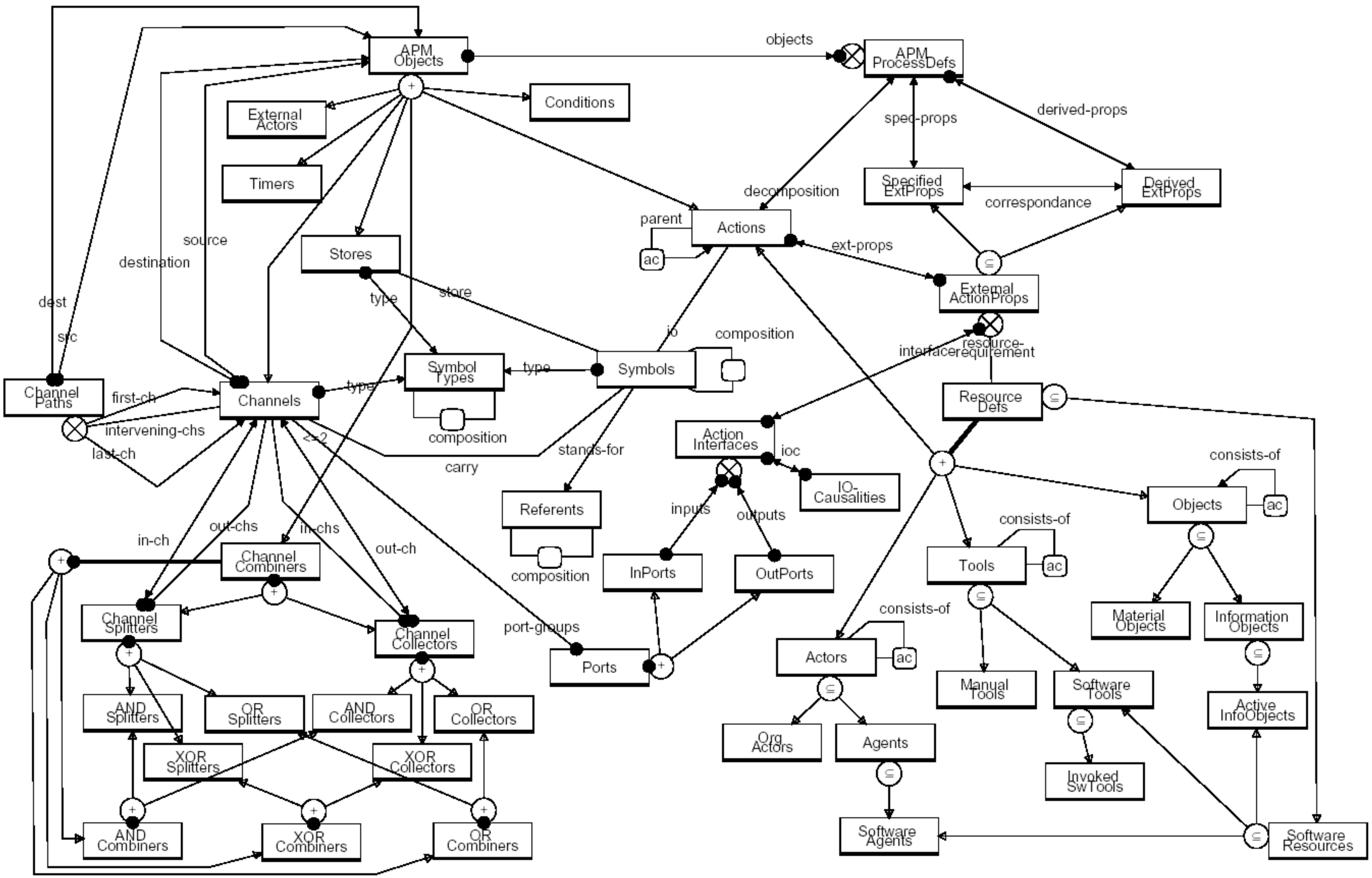


Figure 5.78: APM metamodel, toplevel view

Concepts on different levels of abstraction must be formally defined relative to each other

- **Actions** indirectly have **Symbols** as input / output; denoted by the *io* relation which can be derived through other relationships as follows:
- *io* = *in* ∪ *out*
- *in* = *carry* **o** *port-groups* **o** *inputs* **o** *interface* **o** *ext-props*
- *out* = *carry* **o** *port-groups* **o** *outputs* **o** *interface* **o** *ext-props*

where **o** denotes composition of relations

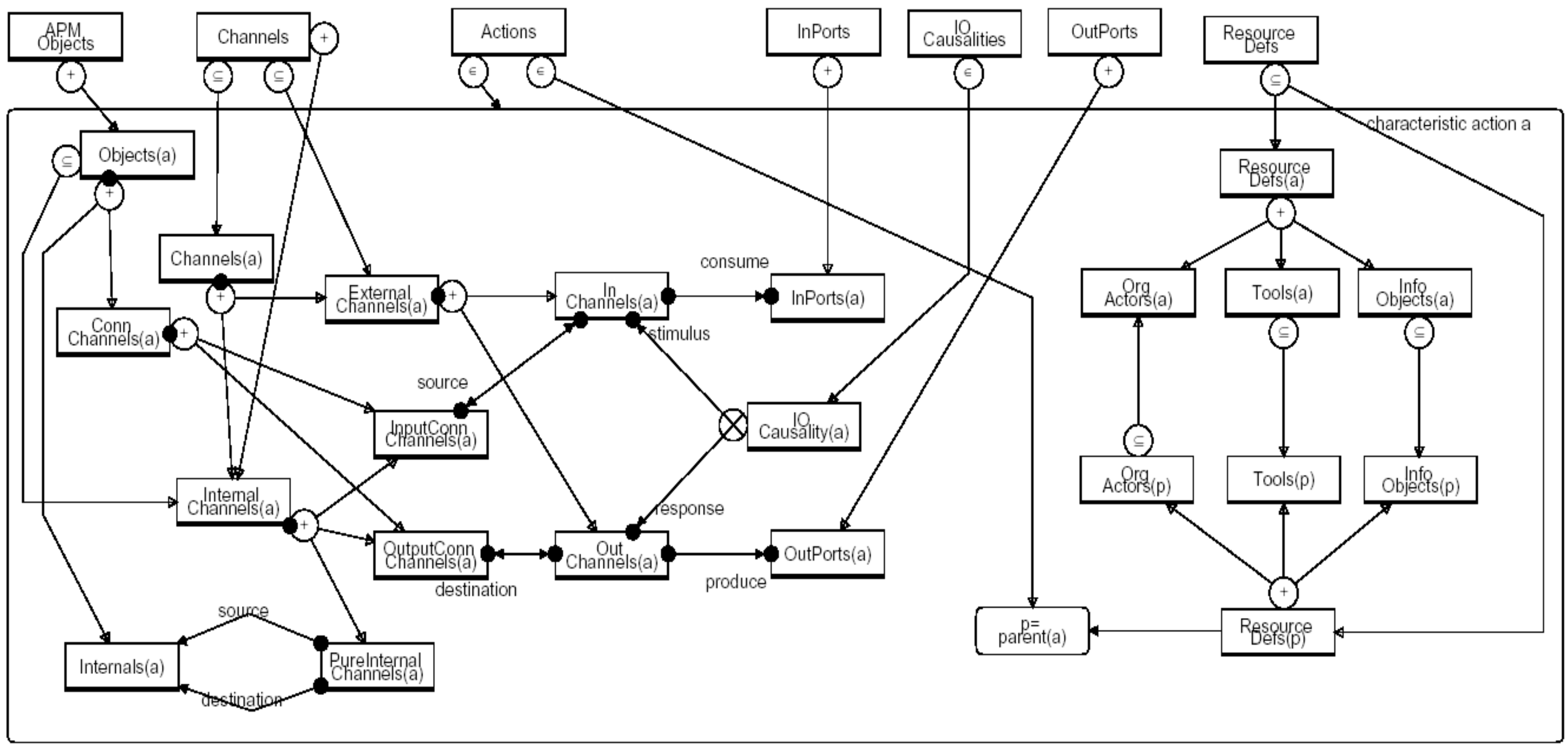
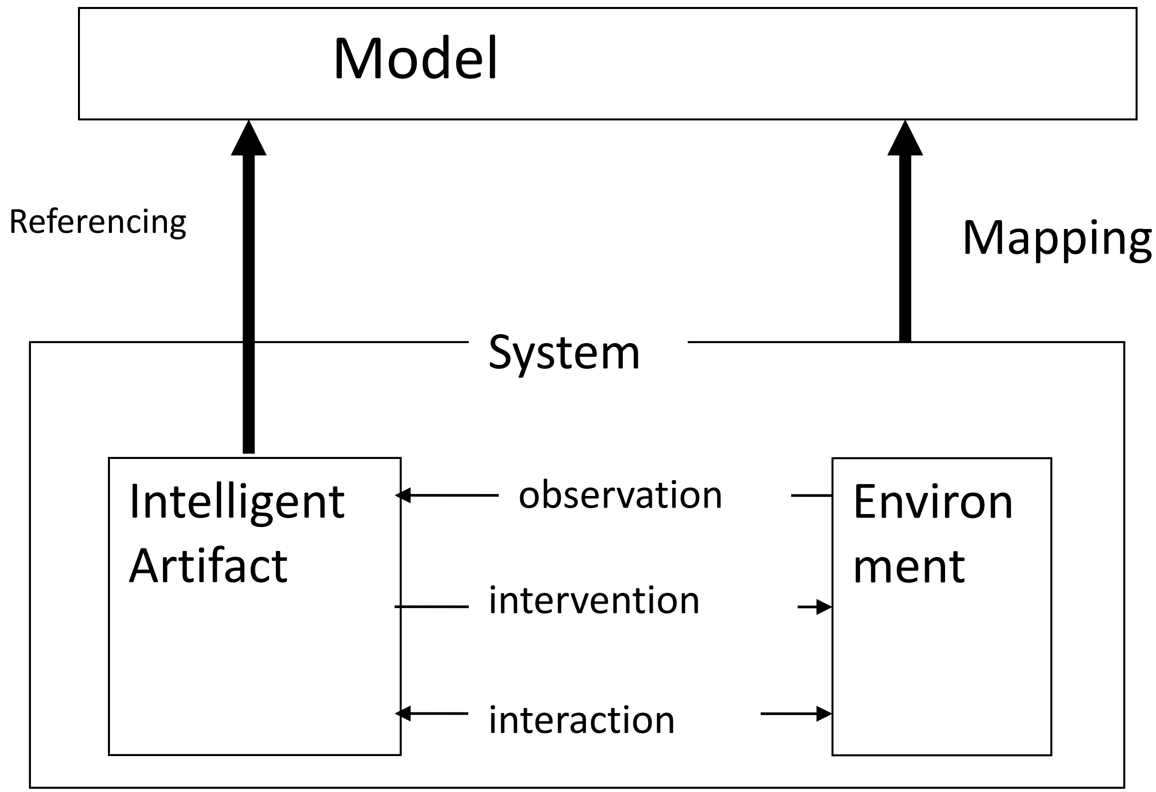


Figure 5.79: APM metamodel, view of characteristic Action element

The “total system”



Knowledge modeling

Search for knowledge represented as models

vs

Search for documents containing knowledge

The Triangle of Meaning

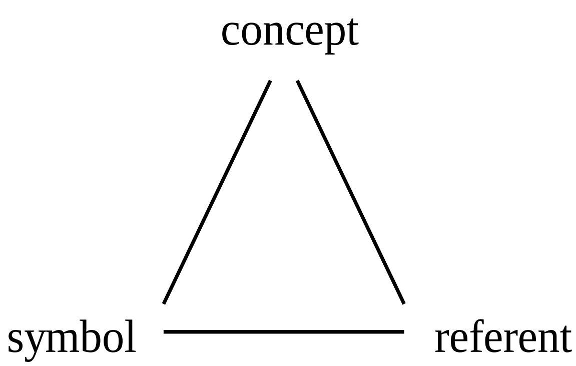


Figure 1 Ogden's triangle

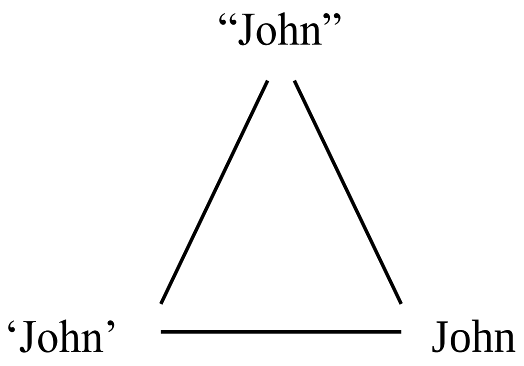
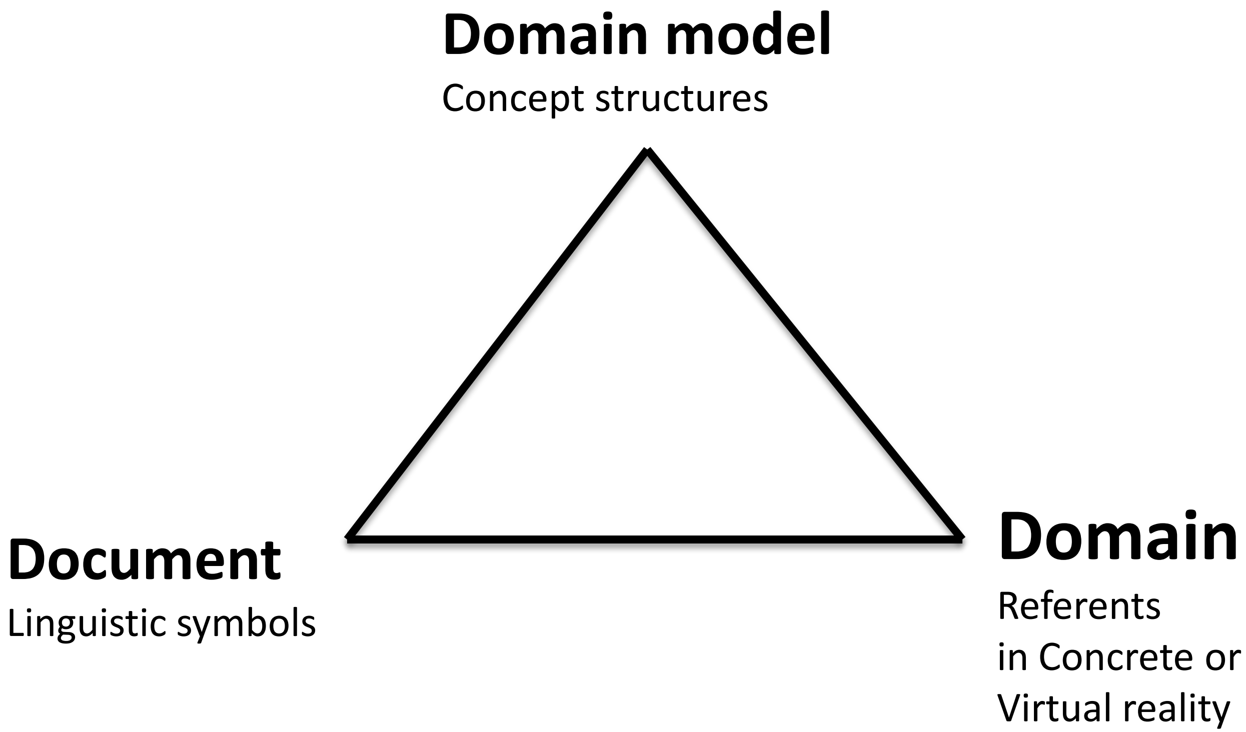


Figure 2 Term, concept and referent

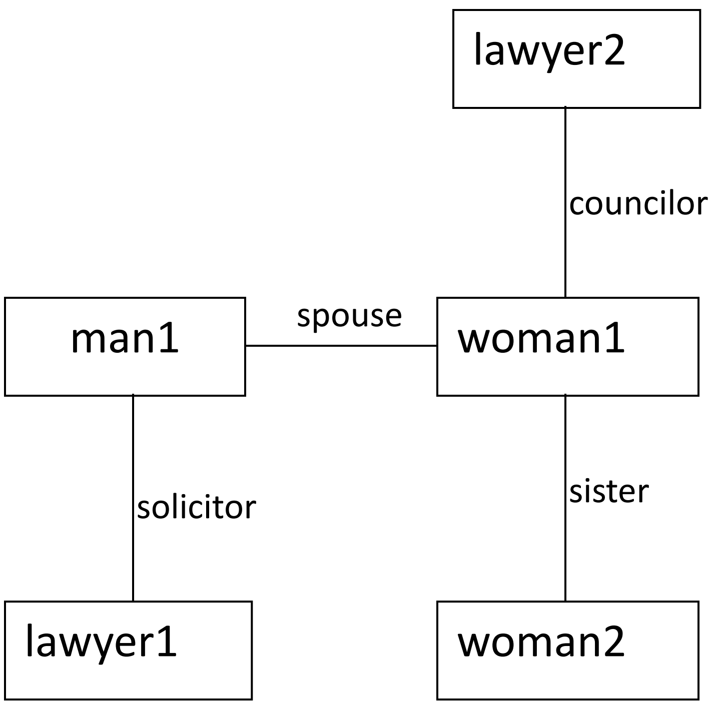
Document, domain, domain model



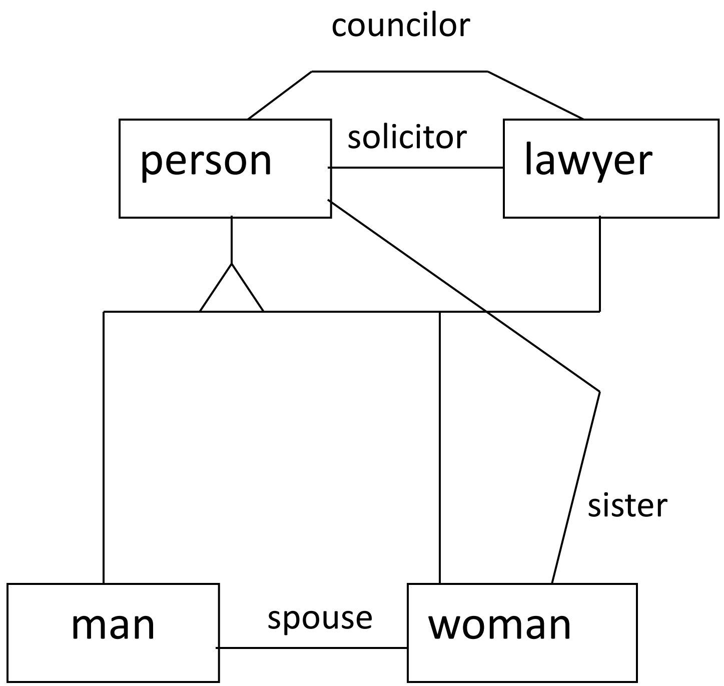
Modeling Ontology

- Document
  - Character
  - Word
  - Phrase
  - Thesaurus (for categorization)
- Domain
  - Concepts
  - Facts
  - Models

Concept model example

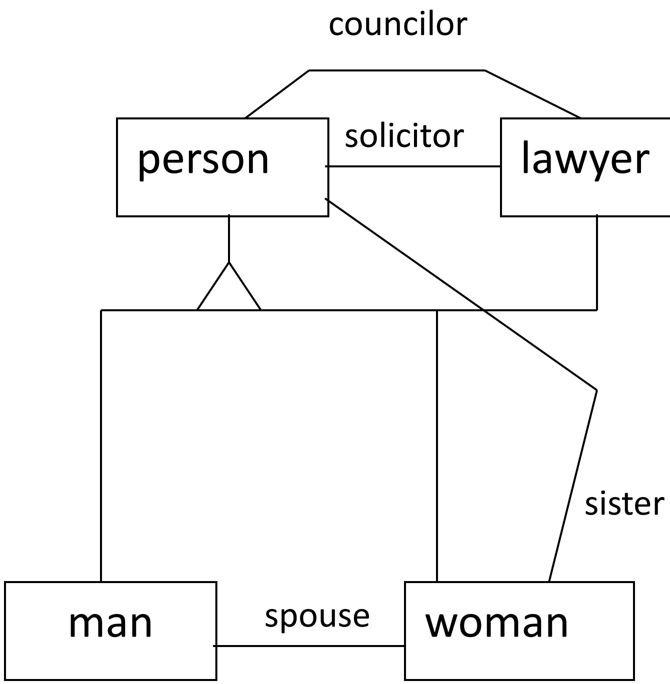


instance model



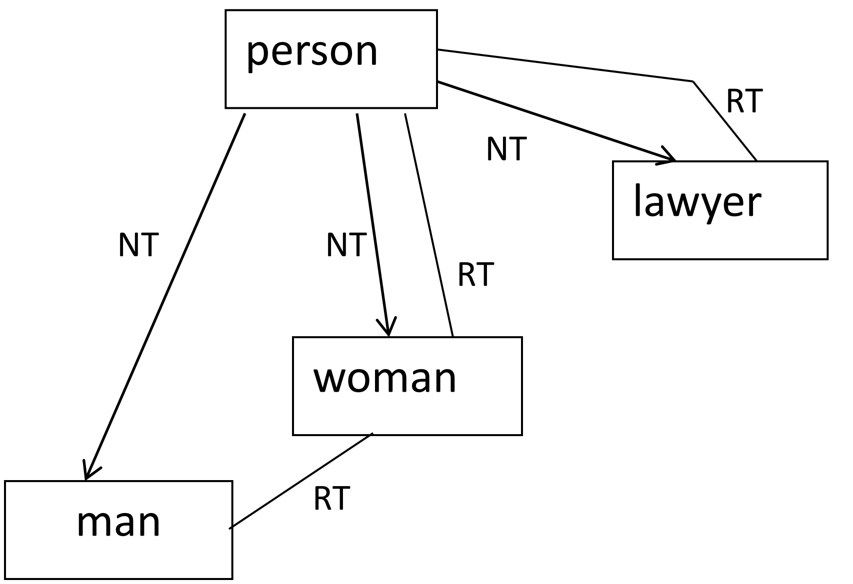
Class/relation model

“Rich” vs. “lightweight” conceptual model



“rich” conceptual model

Legend: NT - narrower term
RT - related term



“light-weight” model

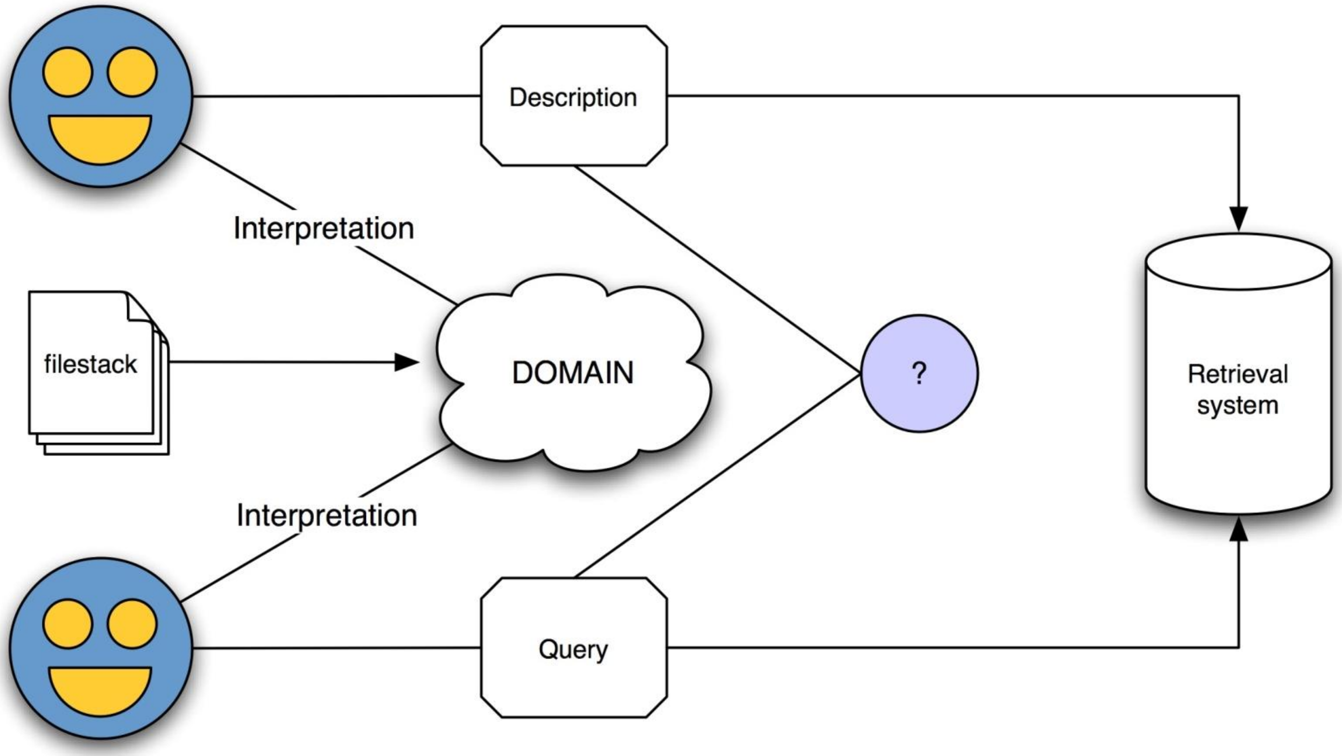
Knowledge retrieval?

- Contemporary knowledge search is based on textual search for relevant documents
- Direct knowledge search must be based on formal domain models
- Is this possible???

Semantic retrieval?

- Semantic is an issue among humans, not computers
- Retrieval is a “communication problem”
- Conceptual modeling enable:
  - explicit, visual representation of domains
  - user readable and applicable domain models
- Linguistic analysis
  - Bridges the gap between abstract model concepts and document texts
  - May support parts of the modeling and retrieval process

The language problem in IR



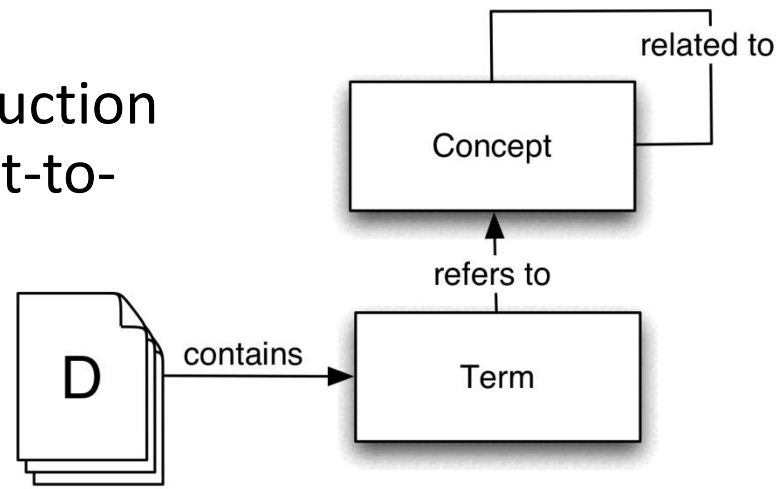
Example
Search for “Definition of CSCW”

- Find
  - no documents containing the term ‘CSCW’
  - 1M+ documents containing the term ‘CSCW’
- Find
  - no explicit definition of the term
  - a number of different definitions or different contexts where the term is applied

A model based approach

- Cooperative domain modelling
  - Explicit definition of concepts
  - Directly applicable in a tool for document description and retrieval

- Linguistic analysis
  - Domain model input and construction of Domain model lexicon for text-to-model matching



Concepts and data

- How do we relate concepts and data?
- Concepts are the units of thinking
- Data are the units of computer storage of information
- How do we relate conceptual models and data models?

Example: weights of people

- Assume that we want to store information about the weight of people on the earth, of the Americans, the British and the Norwegians
- All people are bodies, so all people have weights
- Weights must be represented by numerical values of predetermined accuracy and scale
- The numerical values of individual weight observations are stored in data collections

Ontology of conceptual modeling (scientific style)

- individual concepts
- class concepts
- relation concepts
- quantitative concepts
- specific versus generic concepts
- (linguistic versus non-linguistic concepts)

The quantitative concept

- A quantitative concept is a function $q: UoD \times S \to D$
  - where UoD is the set of referents,
  - S is the set of scales,
  - D is the set of linguistic units (the possible values),
- The set of quantitative concepts is $Q = \{ q \mid q: UoD \times S \to D \}$.

First, the modeling in the domain:

C-concept Body, Person, American, British, Norwegian;
Q-concept weight: Body × Scale × Accuracy→ positive real;
Body includes Person;
Person includes American, British, and Norwegian;

Then we have to introduce appropriate data types:

datatype real wA, real wB, real wN, integer PNO;
Q-concept label: Norwegian ↔ PNO;
Q-concept weight: American × (lbs, d.1d) → wA
Q-concept weight: British × (stone, d.3d) → wB;
Q-concept weight: Norwegian × (kg, d.2d) → wN;

We may now define appropriate information sets:

informationset A-weight, B-weight, N-weight;
A-weight: foreach American (local label, weight);
B-weight: foreach British (local label, weight);
N-weight: foreach Norwegian (PNO, weight);

The concept of weight

- Weight w is the force that a body of mass m exerts on a surface of another body with mass M and radius R
- w=m x G x M/R**2
- All bodies have weights relative to other bodies, depending on masses and distance (R)
- E.g., w(h)=w0 x (1 – 2h/R)
  - w(h) is weight at altitude h
  - w0 is weight at sea level
- Weight is a function w(m,M,R) and simplifications can be applied depending on circumstances

Where does this leave us with respect to

- Gap between concept and data?
- Relation between data, concept and meaning?
- Conceptual models based on shared concepts, e.g., weight, distance, mass, location?
- Overlapping conceptual models?
- Shared understanding of reality?
- Communication without distortion?

Thank you for listening!