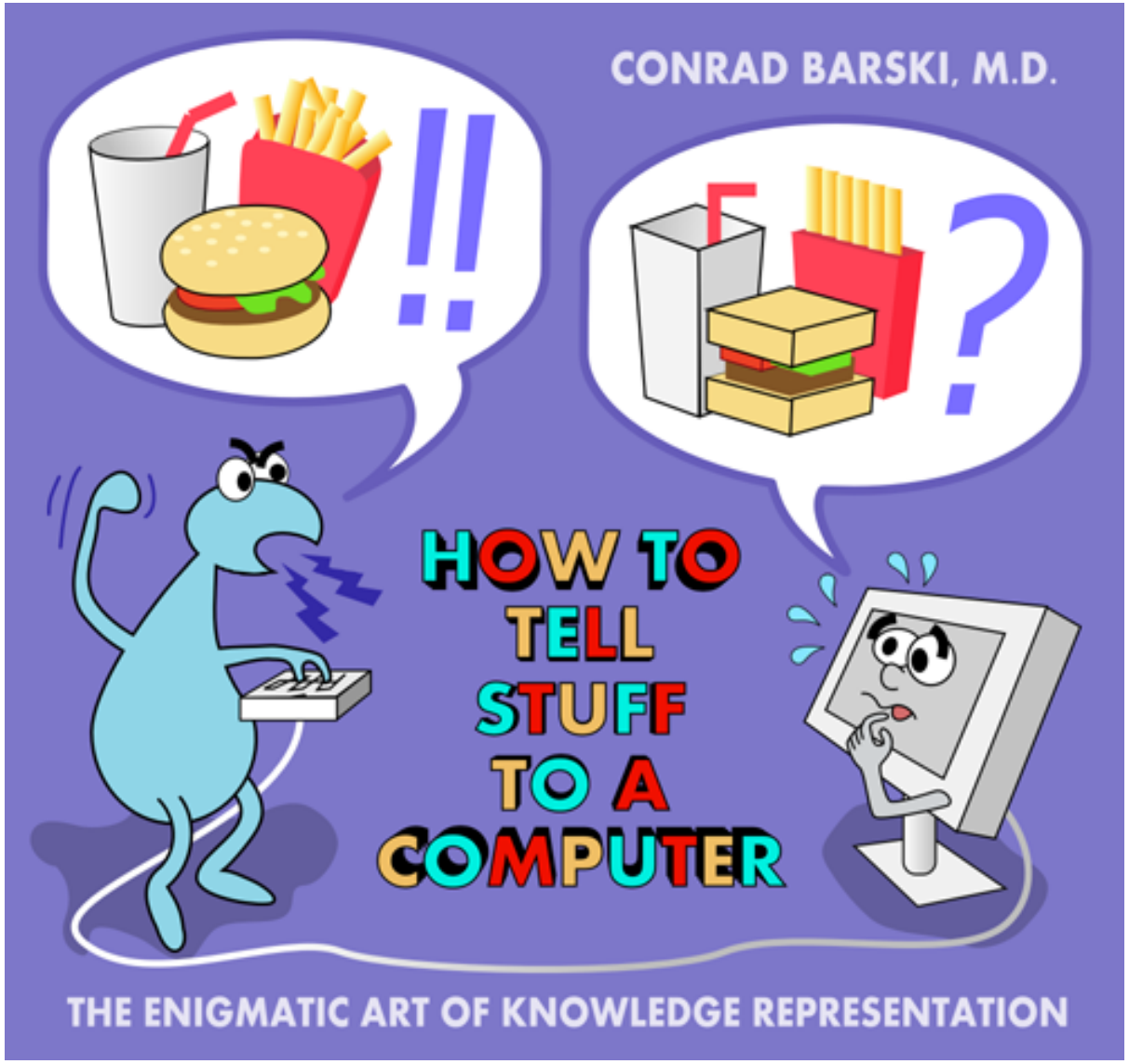
Knowledge Representation



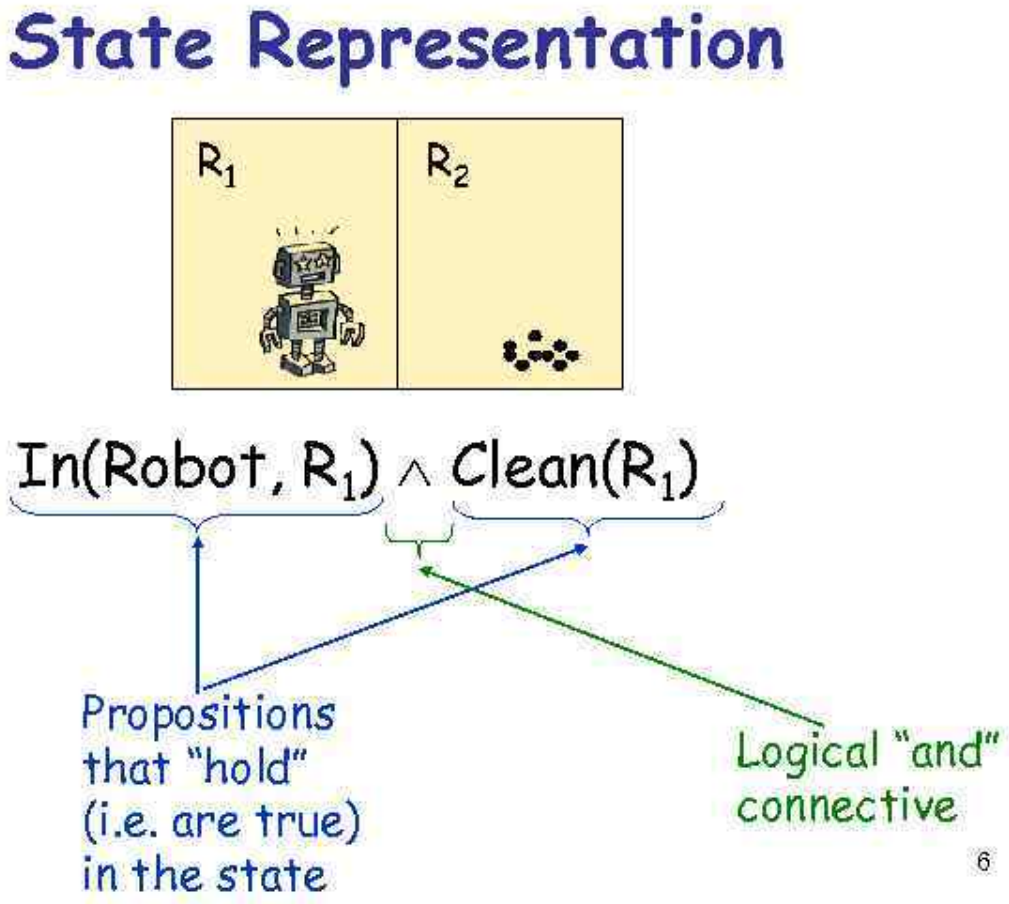
Knowledge Representation: WHAT?

- To express knowledge that can be handled by a computer
  - Symbolic (logic-based, semantic networks, knowledge graphs, ontologies, etc)
  - Numerical (attribute-value, matrices, markov models, gaussian models etc)

Knowledge Representation: WHY?

- to formalize knowledge, two goals:
  - be more precise and non ambiguous
  - to make it "computable"
- to be able to express relations
- to handle categorical data
- to be able to perform **sound reasoning**
- to support learning new knowledge

Knowledge Representation: Example



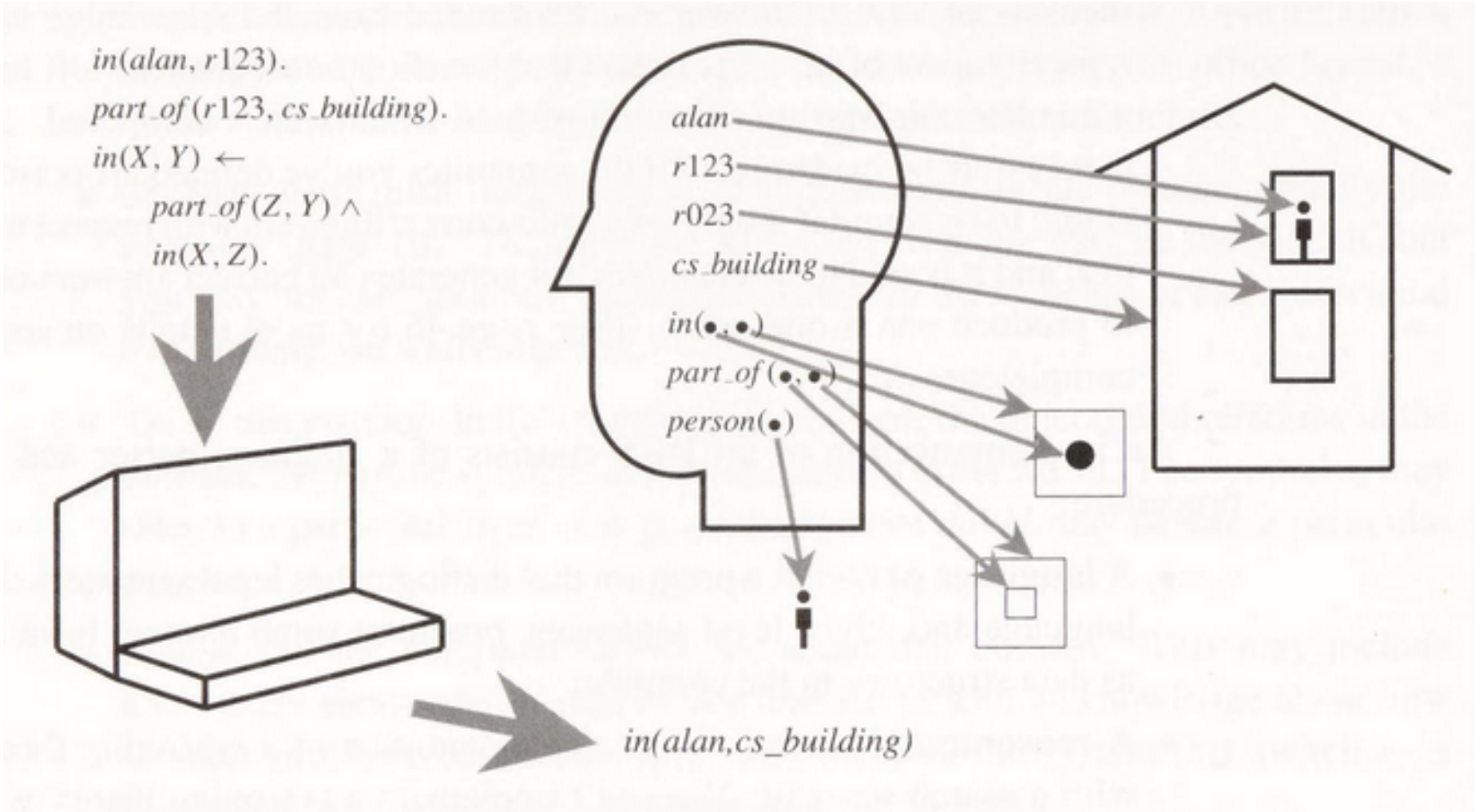
https://present5.com/action-planning-where-logic-based-representation-of-knowledge-makes/

Knowledge Representation

Refs to KR used for these slides:

- chapters 7, 8 e 12, Artificial Intelligence: a Modern Approach, 3rd ed., by Stuart Russell and Peter Norvig
- part II, Artificial Intelligence, 2nd ed., by Elaine Rich and Kevin Knight
- What is a Knowledge Representation?
  `https://groups.csail.mit.edu/medg/ftp/psz/k-rep.html`
- The Cyc Language: `http://www.opencyc.org/doc`
- Prolog: `http://www.dcc.fc.up.pt/~vsc/Yap/`

Knowledge Representation



Exact Knowledge

- **positive:** "Indivíduos sob estresse têm infarte"
- **negative**: "Indivíduos que não estão sob estresse não têm infarte"
- **unknown**: "Mulheres podem ou não ter infarte"

Uncertainty

- **positive:** "Indivíduos sob estresse têm 70% de probabilidade de terem um infarte"
- **negative**: "Indivíduos que não estão sob estresse têm 70% de probabilidade de não sofrerem um infarte"
- **unknown**: "Mulheres têm 10% de probabilidade de sofrer um infarte" (prevalência)

Representation

- Knowledge x Data?
- **Knowledge:** "symbolic representation of aspects of some universe of speech"

Examples of "knowledge"

- José is an employee at UP
- All UP employees have salaries above 25,000 euros (:-)
- All UP employees know they must have a good lifestyle
- José does not think he has a good lifestyle
- All who know José does not think he has a good lifestyle are disappointed

Representation

- **Data:** "symbolic representation of **simple** aspects of some universe of speech"
- special case of "knoweledge"

Exemplos de "dados"

- José is married to Maria
- José is an employee at UP
- The average salary at UP is 25,000 euros

Representation

- Knowledge Representation: to express **knowledge** in a way that can be handled by the computer.

Differents formalisms

Natural Language	Rules
Databases	Decision Trees
Frames	Logic
Scripts	Ontologies
Semantic Networks	Causal Networks
Genetic Algorithms	Neural Networks
Constraints	Markov models
Programming Languages and Data Structures	Object Oriented ...

Representation using Natural Language

Texto Clínico

"Forwarded because of asymmetric density in the ULQ of the left breast. This change has been around since 2005 but the ultrasound evaluation of the exterior suggests the need for biopsy. Breast exam with palpable change with about 30 mm in the ULQ of the left breast."

Disavantages:

ambiguous, redundant, weak structure, sintax and semantics not well understood.

Representation using Databases

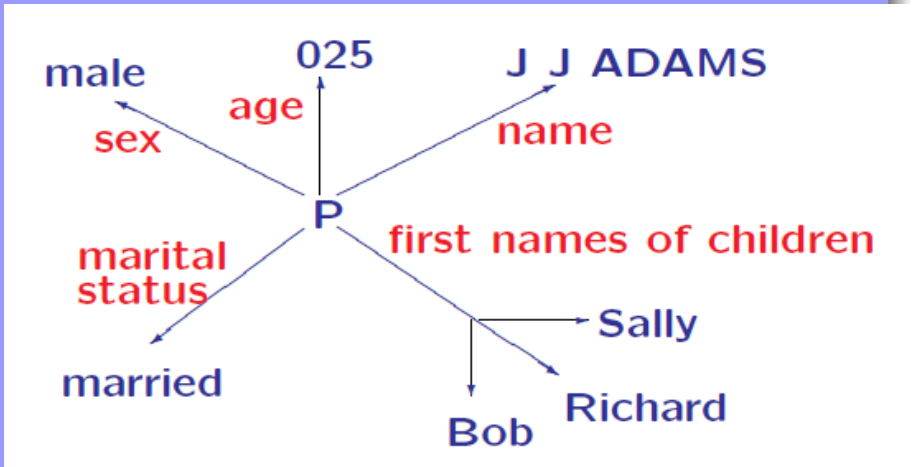
Database

```
person
record = { name : max 20 characters
           age  : 3 digits in range 000-120
           sex  : male or female
           marital status : married, bachelor,
                            spinster, divorced,
                            widowed, or engaged
           first names of children : up to 10 names
                                     each max 15 characters
         }
```

Representation in Databases: an instance

Instance

J. J. ADAMS
025
male
married
Sally
Richard
Bob



Discussion

- only simple aspects can be represented (data)
- entity and relations
- **Reasoning = lookup**

Representation in a single table

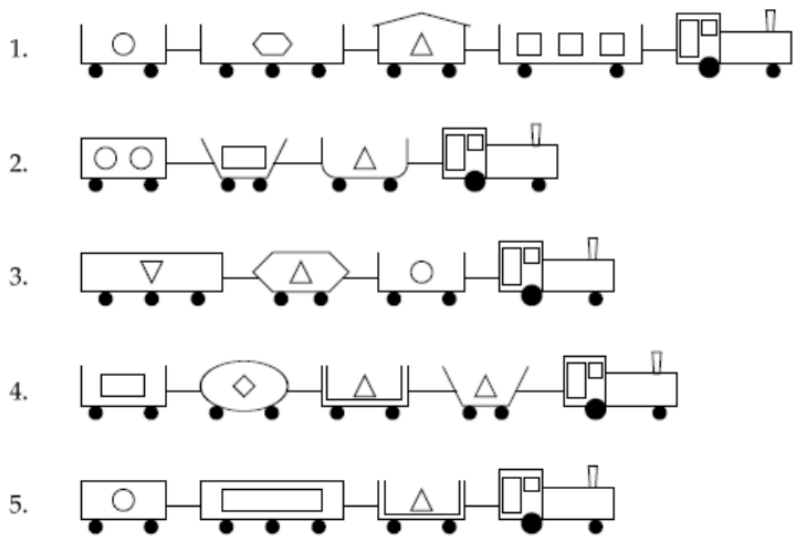
Most learners use this kind of representation: single flat table

Usual: aggregate data in a single table!

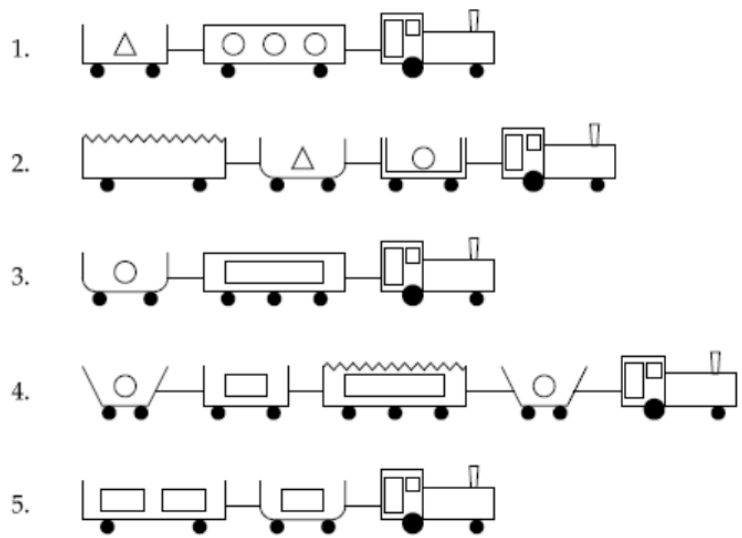
Patient	Location	Size	Date	Calcifications
P1	C	0.1	20050403	F, A
P1	C	0.2	20060412	F
P1	9	0.1	20060412	A
P2	12	0.3	20050415	M
...	...	...	...	...

Exercise

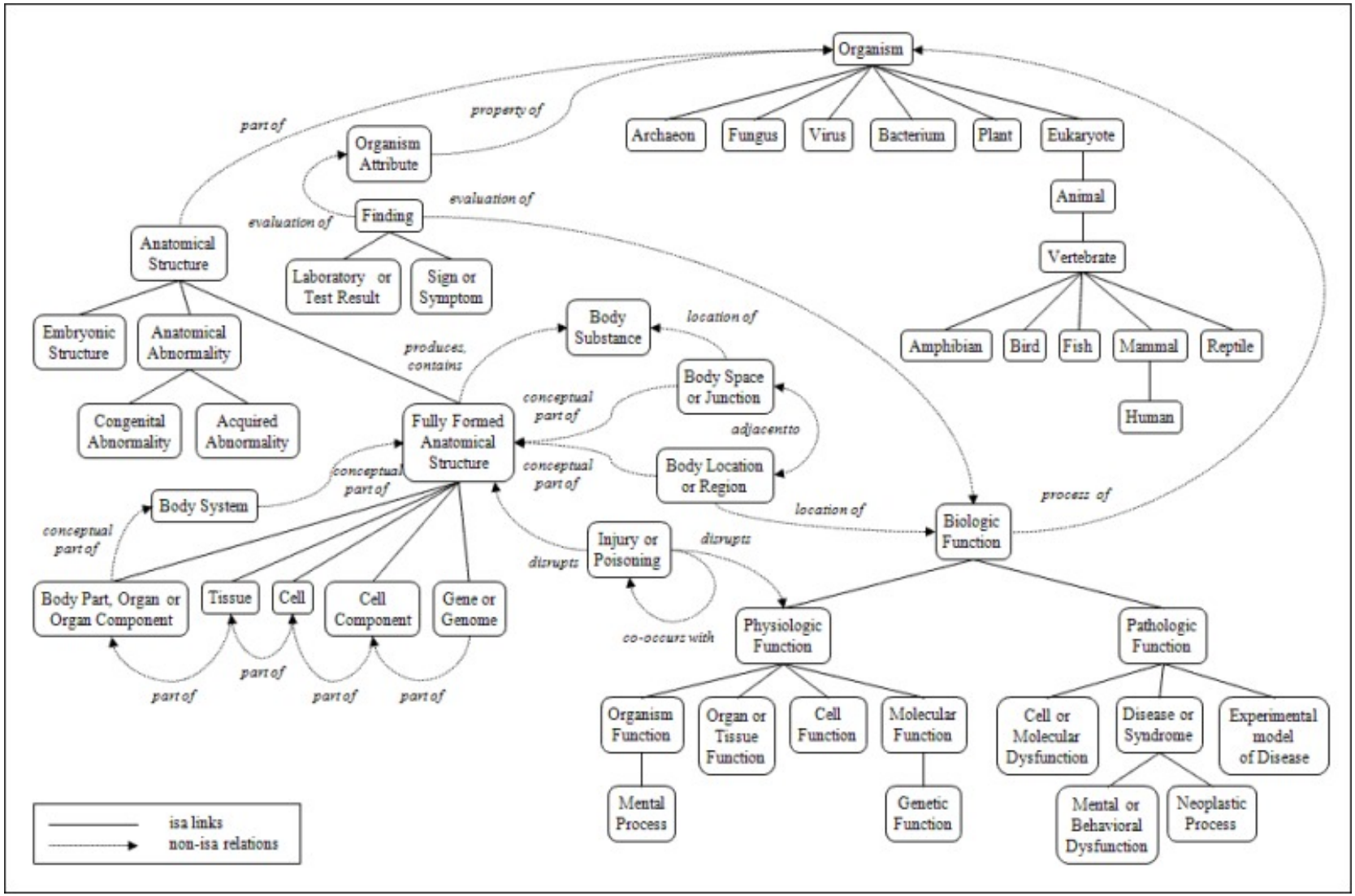
1. TRAINS GOING EAST



2. TRAINS GOING WEST

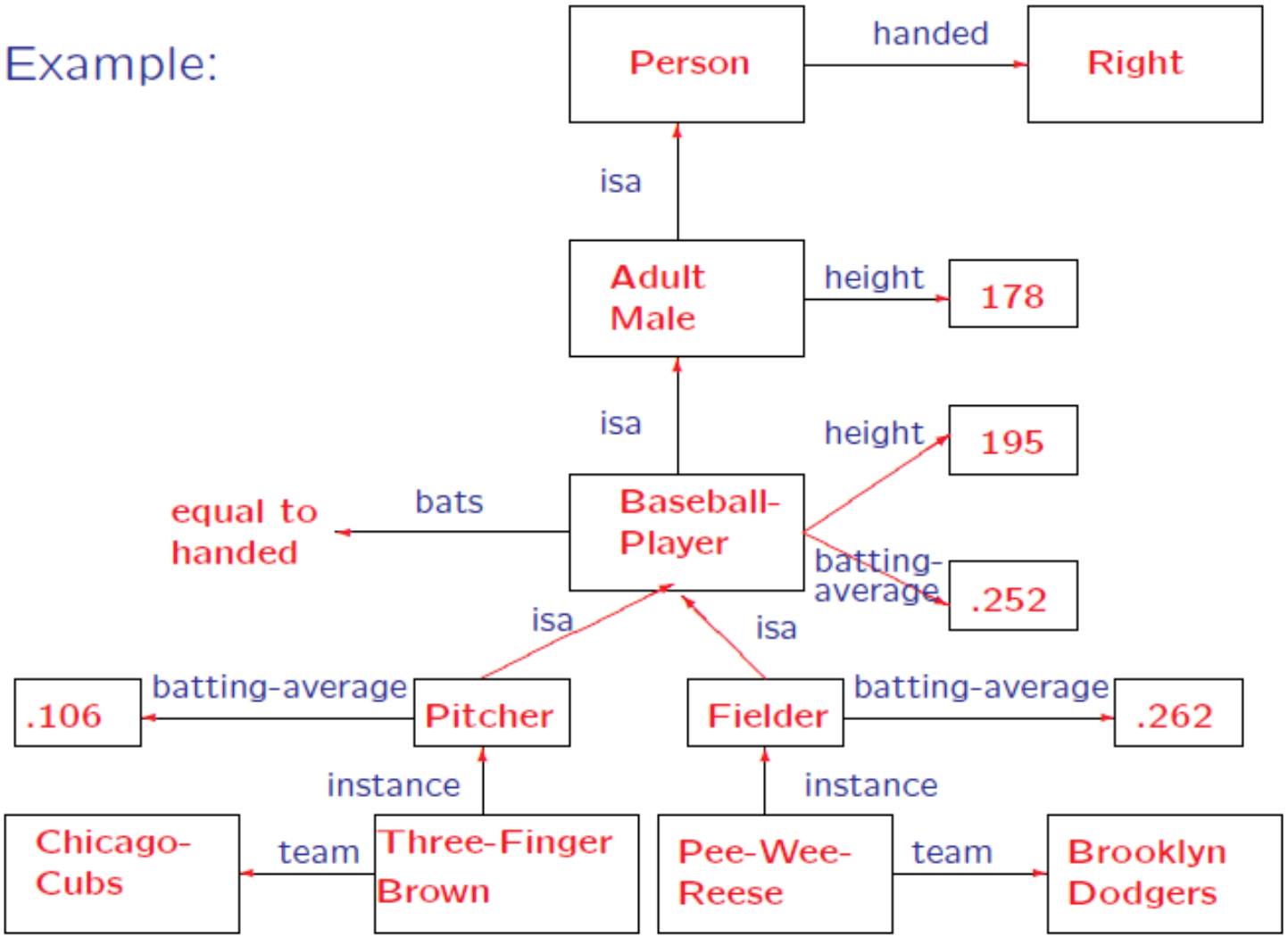


Representation in Semantic Networks



http://umls.it.ilstu.edu/umls.html

Representation in Semantic Network



https://www.cs.bham.ac.uk/~mmk/Teaching/AI/l6.html

Properties of Semantic Networks

- allow structuring knowledge
- use "default" values
- clear sintax, but semantic needs to be worked out

Example based in the Cyc language

"Donald Trump belongs to the collection of U.S. presidents"

```
(#$isa #$DonaldTrump #$UnitedStatesPresident)
```

"All trees are plants"

```
(#$genls #$Tree-ThePlant #$Plant)
```

"Paris is the capital of France."

```
(#$capitalCity  #$France  #$Paris)
```

"if OBJ is an instance of the collection SUBSET and SUBSET is a subcollection of SUPERSET, then OBJ is an instance of the collection SUPERSET".

```
(#$implies
   (#$and
     (#$isa ?OBJ ?SUBSET)
     (#$genls ?SUBSET ?SUPERSET))
   (#$isa ?OBJ ?SUPERSET))
```

Frames

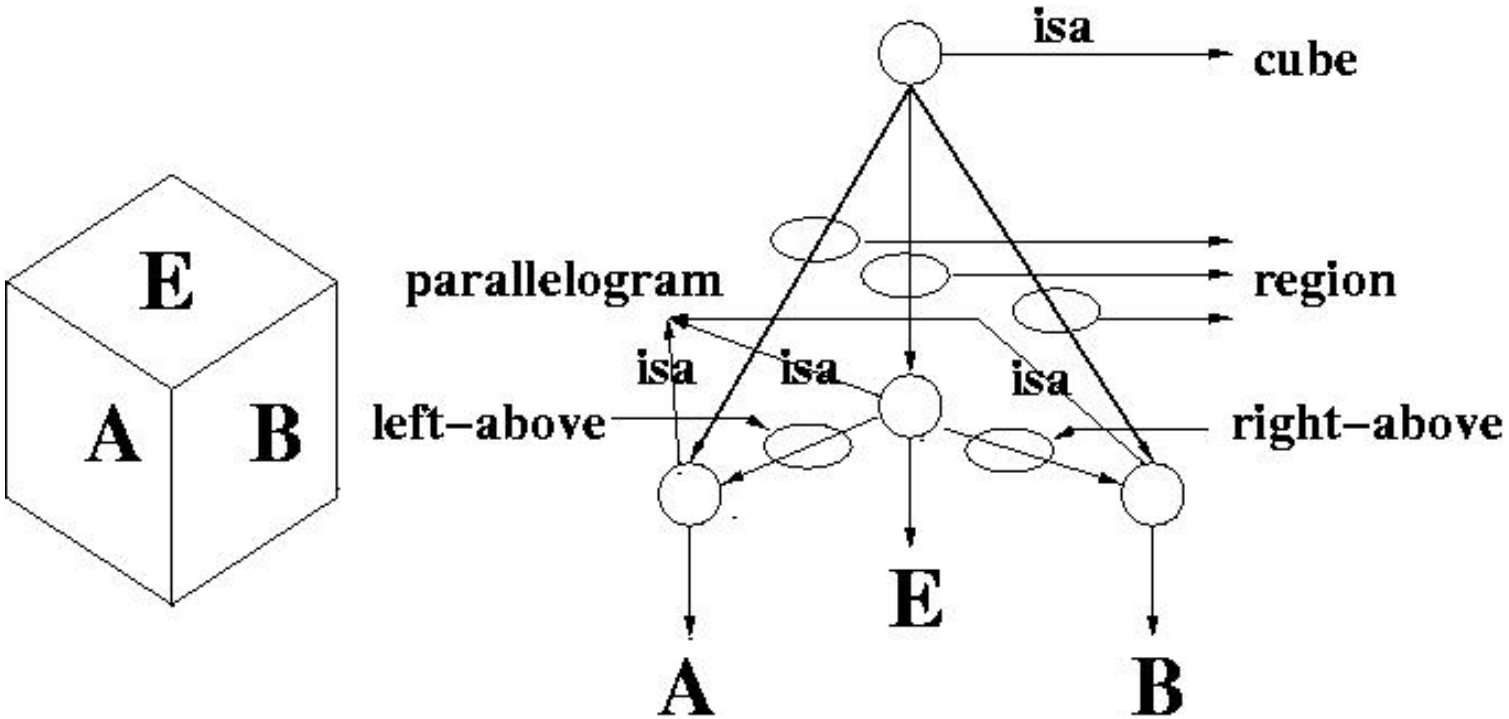
- A Frame is a collection of slots whose contents can be a value or a pointer to another Frame. (how is this different from a relational database??)

Birthday party	
Dress:	sports social
Gift:	must please the birthday person must be bought and wrapped
Games:	hide-and-seek put the tail on the donkey
Decoration:	baloons, surprise bags, crepe paper
Menu:	Cake, Ice cream, Soft drink, hot dog
Cake:	light up candles, blow candles, make a wish, sing happy birthday
Ice cream:	napolitan

Frames

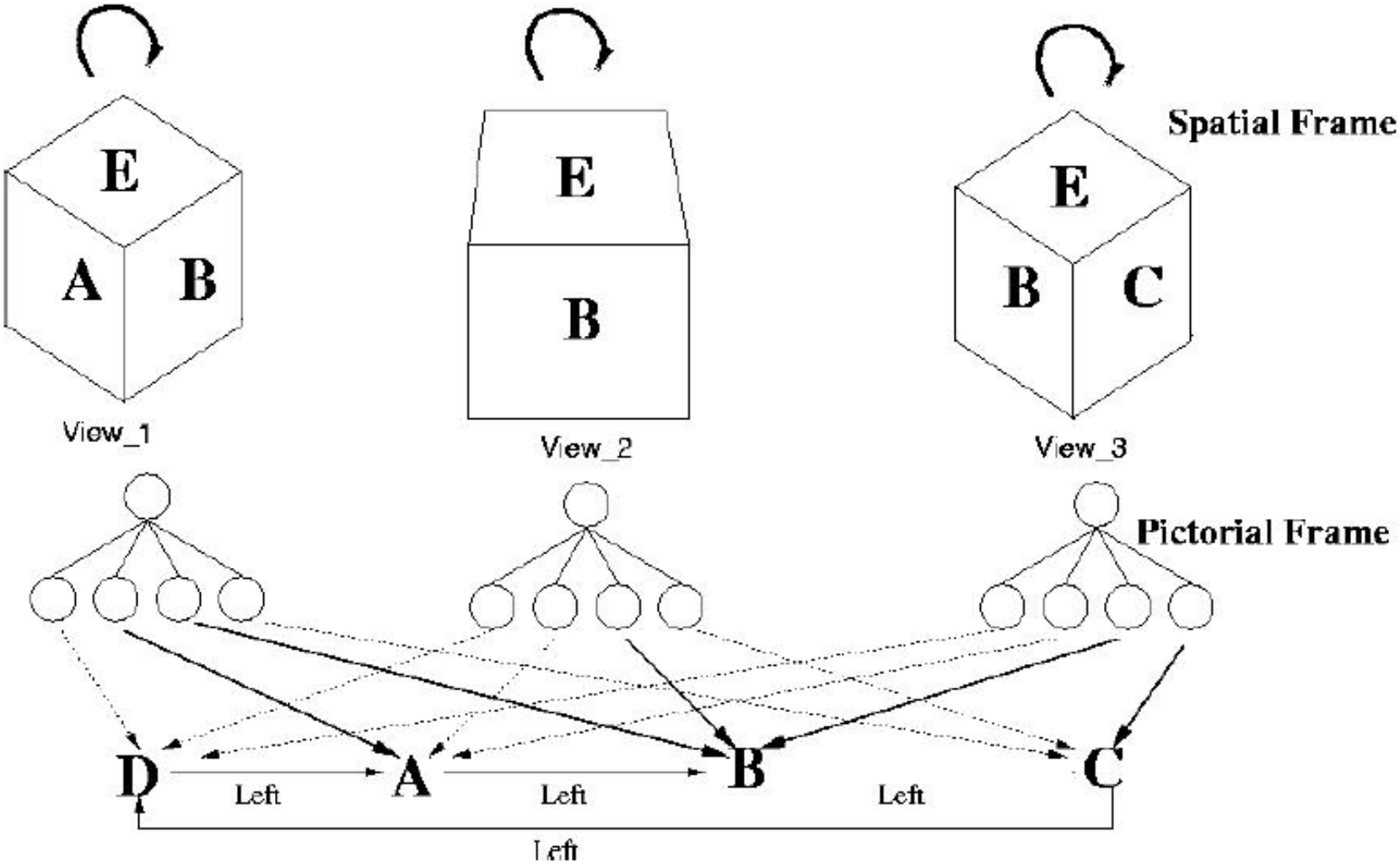
- A frame is a collection of questions to be answered about a hypothetical situation: specify questions and methods.
  - What caused (agent)?
  - What is the purpose (intention)?
  - Whar are the consequences (effects)?
  - Who is affected (receptor)?
  - How is it performed (instruments/methods)?

Exemplo



Object composed by relations.

Example



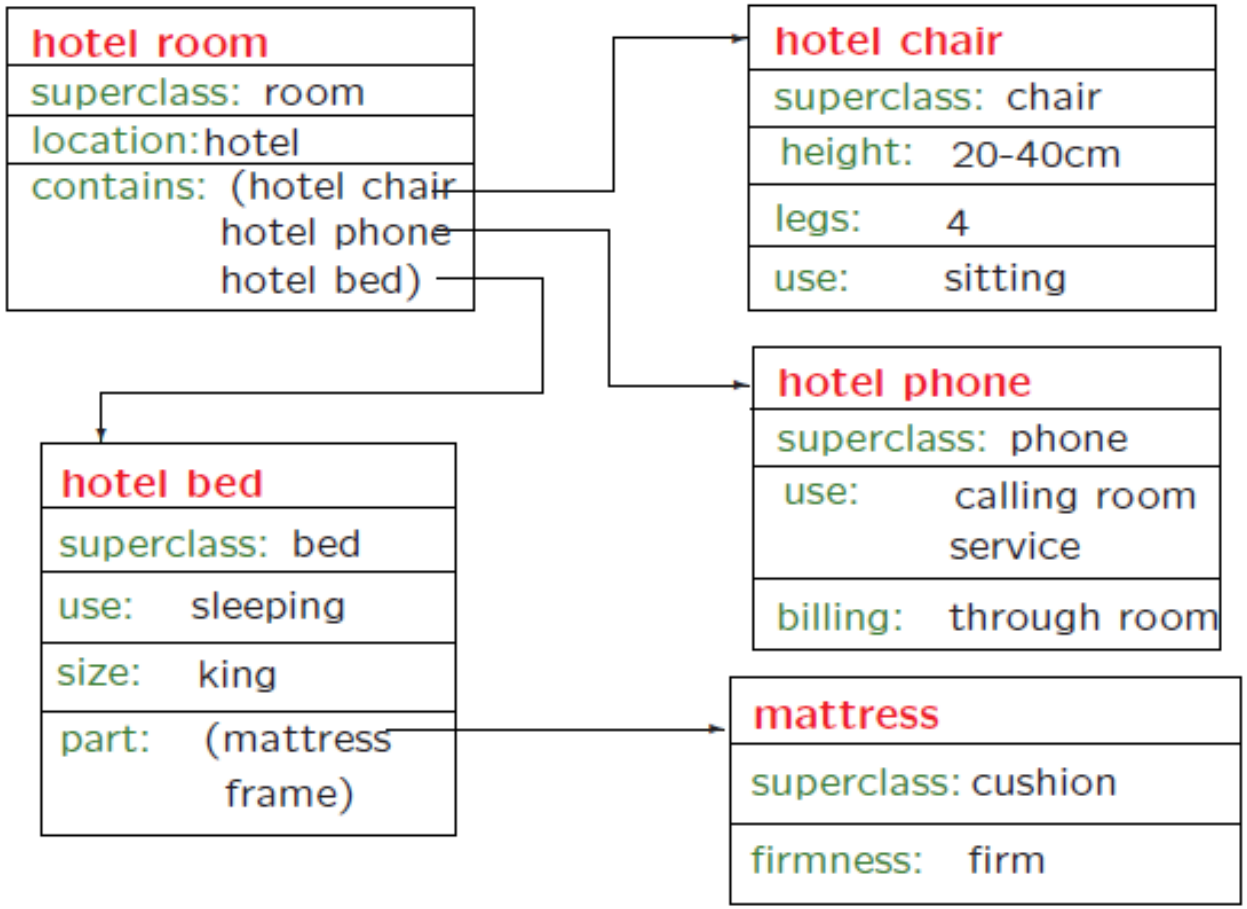
Differents aspects of a cube.

Frame for and aspect of a cube

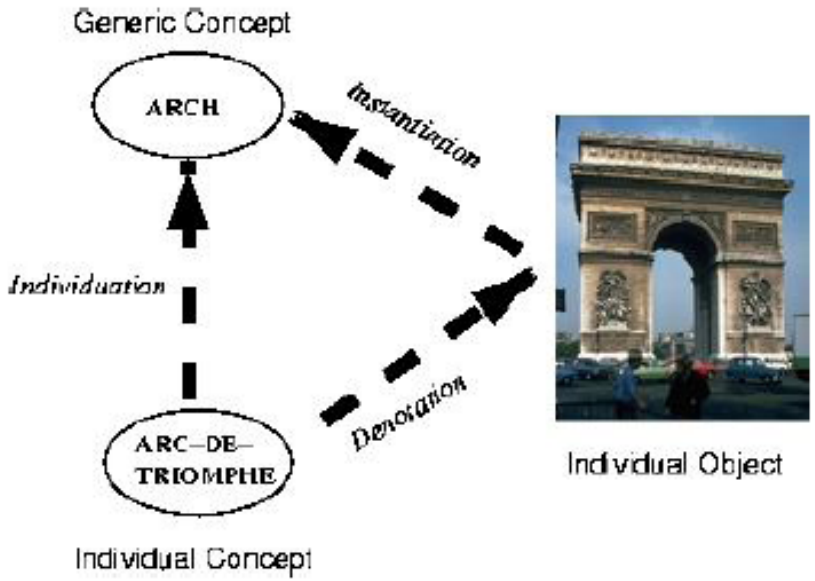
An aspect of a cube using a frame representation

View-of-a-Cube		
Slot	Filler	Constraint
Name	View_1	
region_of	A	parallelogram & visible
region_of	B	parallelogram & visible
region_of	C	parallelogram & invisible
region_of	D	parallelogram & invisible
region_of	E	parallelogram(E) & visible & left-above(E,A) & right-above(E,B)

One more example



Important!



- distinguir:
  - concepts (representations) and objects (instances)
  - individual/specific concepts and general concepts

Associating procedures/methods to the representation

Procedimentos

rectangle	
superclass:	polygon
Coordinates:	(0cm,0cm)
length:	5cm
width:	2cm
area:	procedure(z) length(z) * width(z)
perimeter:	procedure(z) 2 * (length(z) + width(z))

Scripts

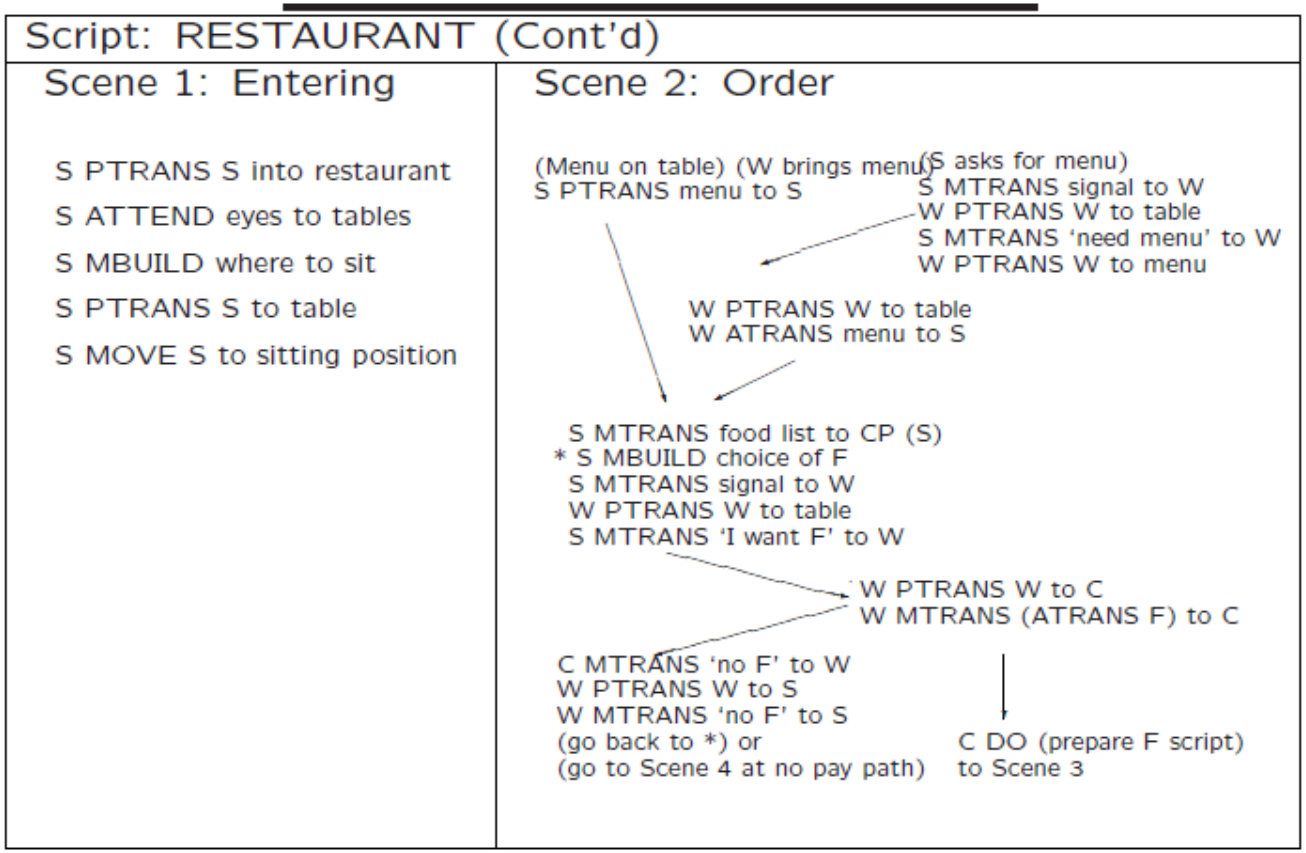
- A script is a structured representation that describes a sequence of events in a given context.
  - Extends frames through an explicit representation of actions and changes of states.
  - Define primitives to describe the universe of speech:
    - PTRANS physical transfer of an object ("go")
    - ATRANS transfer of relations ("give")
    - MTRANS mental transfer ("tell")

Script Example

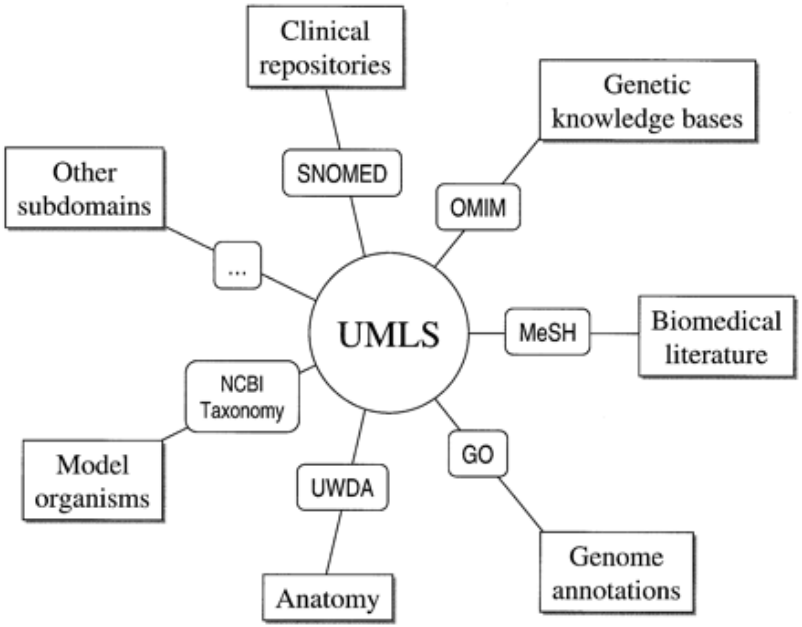
Script for a restaurant

Script: RESTAURANT			
Track:	Coffee Shop	Entry cond.:	S hungry
Props:	Tables		S has money
	Menu		
	F=Food	Results:	S has less money
	Check		O has more money
	Money		S is not hungry
Roles:	S=Customer		
	W=Waiter		
	C=Cook		
	M=Cashier		
	O=Owner		

Script for a restaurant (cont.)



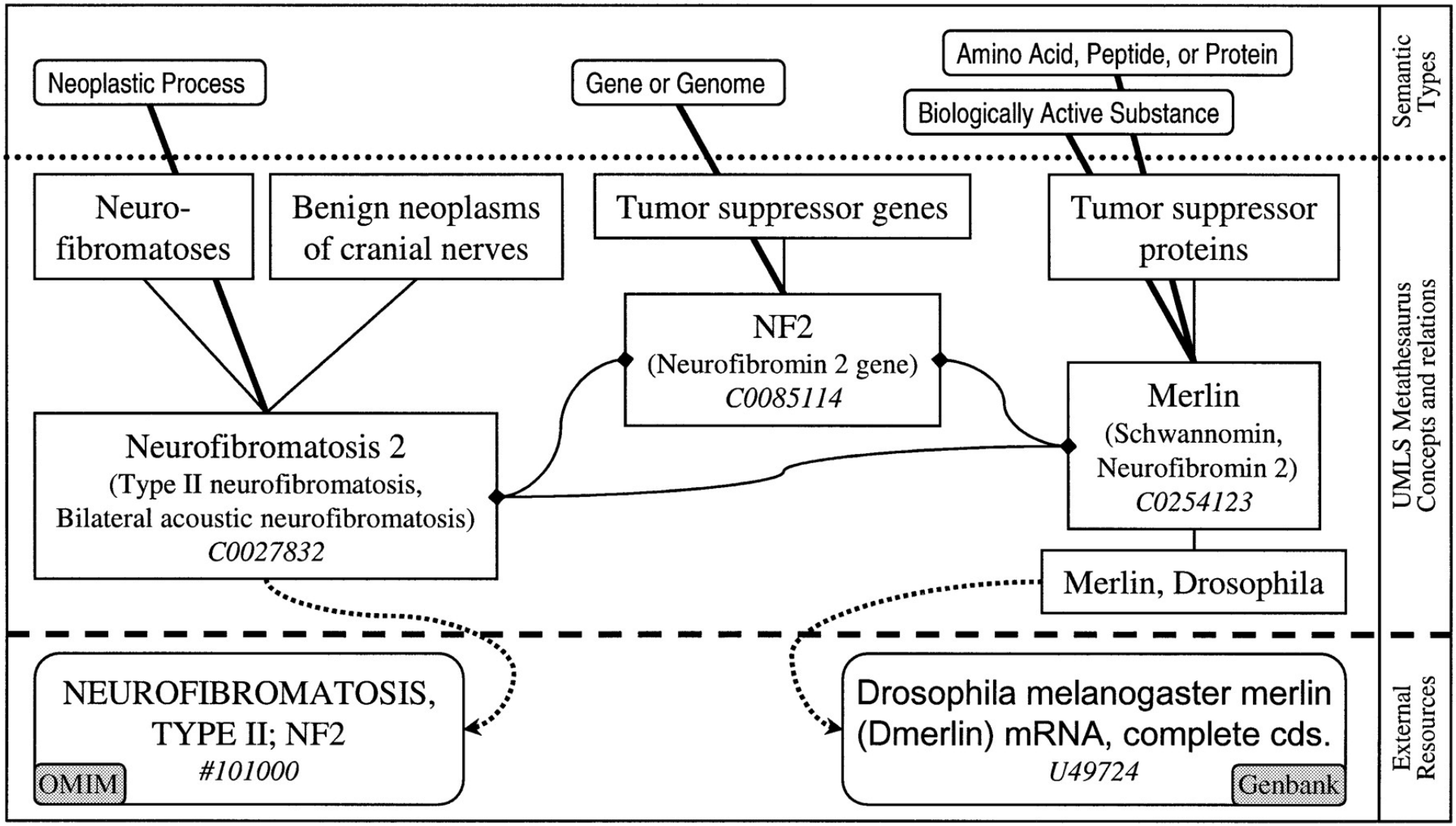
Unified Medical Language System – UMLS



MetaTheasaurus

The major component of the UMLS is the Metathesaurus, a repository of inter related biomedical concepts. The two other knowledge sources in the UMLS are the Semantic Network, providing high level categories used to categorize every Metathesaurus concept, and lexical resources including the SPECIALIST lexicon and programs for generating the lexical variants of biomedical terms. (Nucleic Acids Research, Oxford Journals)

Unified Medical Language Systems – Example



https://academic.oup.com/nar/article/32/suppl_1/D267/2505235

XML

```
<?xml version="1.0" encoding="UTF-8"?>
<Patient id="122">
  <General>
    <Weight-kg>35</Weight-kg>
    <Height-cm>128</Height-cm>
    <Day>05</Day>
    <Month>02</Month>
    <Year>2004</Year>
    <Sex>1</Sex>
    <AuscultationPosition>2</AuscultationPosition>
  </General>
  <SystemicPressure>
    <SystemicPressureMethod>1</SystemicPressureMethod>
    <SystolicSystemicPressure-mmHg>130</SystolicSystemicPressure-mmHg>
    <DiastolicSystemicPressure-mmHg>90</DiastolicSystemicPressure-mmHg>
  </SystemicPressure>
  <PulmonaryPressure>
    <PulmonaryPressureMethod />
    <SystolicPulmonaryPressure-mmHg />
    <DiastolicPulmonaryPressure-mmHg />
    <CatheterismSimultaneousMeasurement />
    <EchocardiograSameConsultation />
  </PulmonaryPressure>
  <Murmur>
    <Cycle>2</Cycle>
  </Murmur>
  <S1>
    <S1Status>1</S1Status>
  </S1>
  <S2>
    <S2Status>1</S2Status>
    <IfAbnormal>0</IfAbnormal>
    <PulmonaryComponent>1</PulmonaryComponent>
```

procedural/imperative or declarative?

- How? Procedural/Imperative
- What? Declarative
- Calculating the sum of elements of an array:

```
   Procedural                    Declarative
   (array is indexed)           (array is a list)
sum = 0                          sum([],0).
for i = 0 to arraysize           sum([H|T],N+1) :- sum(T,
   sum += array[i]
end for
```

Properties of "good" representations

- important objects and their relations are explicit
- express natural constraints
- represent objects and relations together
- omit irrelevant details
- transparent: easy to understand
- complete
- concise
- quick to store and retrieve
- "computable"

Properties of "good" representations

- lexical part determines which symbols can be used
- structural part describes how the symbols can be organized (constraints)
- procedural part specifies procedures that allow to create, modify and ask questions (query) the descriptions
- semantic part establishes a meaning to the description

Properties of "good" representations

For example, in the semantic networks:

- lexical: nodes, links, link labels
- structure: directed graph with labeled edges
- procedural: constructors, readers, writers, erasers (to create and modify the graph)
- semantic: meaning of nodes and edges depends on the domain and application

Propositional Rules

The first rule says "if there has been one previous Hib dose (Hib_prior = 1) and the Hib series is active and the Hib dose 1 was given at over 12 months of age and the Hib2_final parameter set is met (e.g., the minimum ages and wait-interval criteria are satisfied), then dose Hib 2 is due, and the parameters in the Hib2_final parameter set apply."

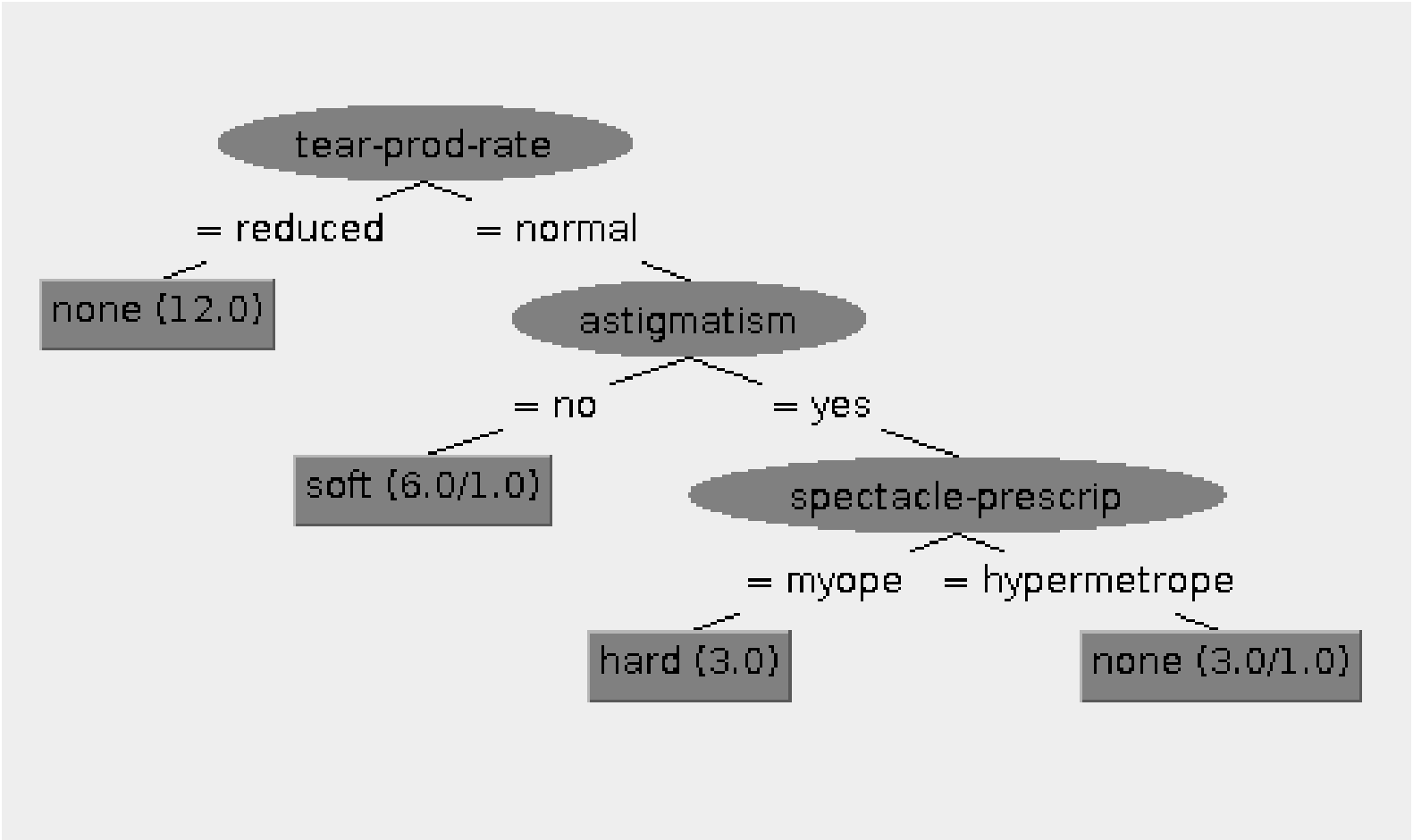
```
if:   Hib.prior = 1 and not Hib_inactive and Hib1_age_in_months >= 12
      and Hib2_final_parameters_met
then: due.Hib2_final

if:   Hib.prior = 1 and not Hib_inactive and Hib1_age_in_months < 12
      and Hib2_parameters_met
then: due.Hib2

if:   Hib.prior = 1 and not Hib_inactive and Hib1_age_in_months >= 12
      and not Hib2_final_parameters_met
then: next.Hib2_final

if:   Hib.prior = 1 and not Hib_inactive and Hib1_age_in_months < 12
      and not Hib2_parameters_met
then: next.Hib2
```

source: Decision Support and Expert Systems in Public Health, in Public Health Informatics and Information Systems, edited by Patrick W. O'Carrol

Decision Trees

- Simple and easy to implement.
- The input is an object described by its properties
- The output can be an answer "yes" or "no"
- Such trees represent boolean functions (but can be more general)

Decision Trees: Example



Logic

- Languages:
  - **sintax**: describe possible configurations of the language that constitute **valid** sentences.
  - **semantic**: determines the meaning of each sentence.
- example: $x > y$,
  - sintax: if x is a number and y is a number, then $x > y$ is a sentence over numbers.
  - semantic: if $x > y$ return true, otherwise return false.

Logic

- Relations represented as tuples:

```
margin(massID,spiculated).
```

- Relations may depend on other relations:

```
IF margin(massID, spiculated) AND
  size(massID,5) THEN
       malignant(massID).
```

- Symbols: and, or, if-then

Logic

- Representations can be more complex:

```
IF lesion(x) AND connected(y,x) THEN lesion(y)
```

Examples of First-Order Rules

$$\begin{aligned} same\_finding(F_1, F_2) \longleftarrow & MLOView(F_1) \wedge CCView(F_2) \wedge \\ & nipple\_distance(F_1, D_1) \wedge nipple\_distance(F_2, D_2) \wedge \\ & \big(abs(D_1 - D_2) < \epsilon\big) \wedge \\ & side(F_1, left) \wedge side(F_2, left) \wedge \\ & quadrant(F_1, upper\_outer) \wedge quadrant(F_2, upper\_outer) \wedge \\ & massShape(F_1, oval) \wedge massShape(F_2, oval). \end{aligned}$$

$$\begin{aligned} previous\_finding(F_1, F_2) \longleftarrow & mammo(P, F_1) \wedge mammo(P, F_2) \wedge \\ & date(F_1, D_1) \wedge date(F_2, D_2) \wedge \\ & (D_1 < D_2 \vee D_2 < D_1) \end{aligned}$$

This rule relates two findings F_1 and F_2 for the same patient P, separated in time (date of F_1 is before or after the date of F_2). It can be further used to simulate temporal reasoning in the context of other rules such as:

$$\begin{aligned} is\_malignant(A) \longleftarrow & mass(A, present) \wedge previous\_finding(A, B) \wedge \\ & \big(massSize(A) < massSize(B)\big) \wedge calc(B, present) \wedge \\ & previous\_finding(A, C) \wedge calcFineLinear(C, yes) \end{aligned}$$

source: Automated Diagnosis of Breast Cancer on Medical Images, in Foundations of Biomedical Knowledge Representation, edited by Hommersom and Lucas, Springer, 2016

Exercise # 1

Represent the following sentences using first-order logic:

- Example 1
  - All viral diseases are infectious diseases
  - covid-19 is an infectious disease
- Example 2
  - Whoever has acute hepatitis has hyperbilirubinemia
  - If someone has hyperbilirubinemia, this person also has jaundice
- Example 3
  - All viral diseases are infectious diseases
  - Myocardial infarction is not a viral disease

Exercise # 2

Represent the following sentences using first-order logic:

- a) architectural distortion, spiculated masses, and distributed microcalcifications are lesions.
- b) These lesions appear in the breast region.
- c) The breast is a region related with another region, the armpit.
- d) Lesions connected to other lesions are lesions.
- e) Lymphoma can appear in the armpit.
- f) Barbazul is parent of Silver.
- g) descendents and parents are inverse relations.
- h) All mammals have parents.