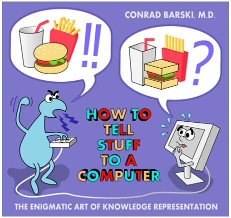
Clinical Decision Support Systems, 23/24

Inês Dutra and Pedro Rodrigues
DCC-FCUP & MEDCIDS-FMUP
ines@dcc.fc.up.pt, pprodrigues@med.up.pt

February 23rd, 2024



Bibliography

- Artificial Intelligence: a Modern Approach, by Stuart Russell and Peter Norvig, 3rd edition, Prentice Hall (chapters 7–9, 12)
- Artificial Intelligence: Foundations of Computational Agents, second edition, Cambridge University Press 2017, by David Poole and Alan Mackworth (chapter 14)
- Artificial Intelligence: a new synthesis, by Nils Nilsson
- Artificial Intelligence, by Elaine Rich and Kevin Knight

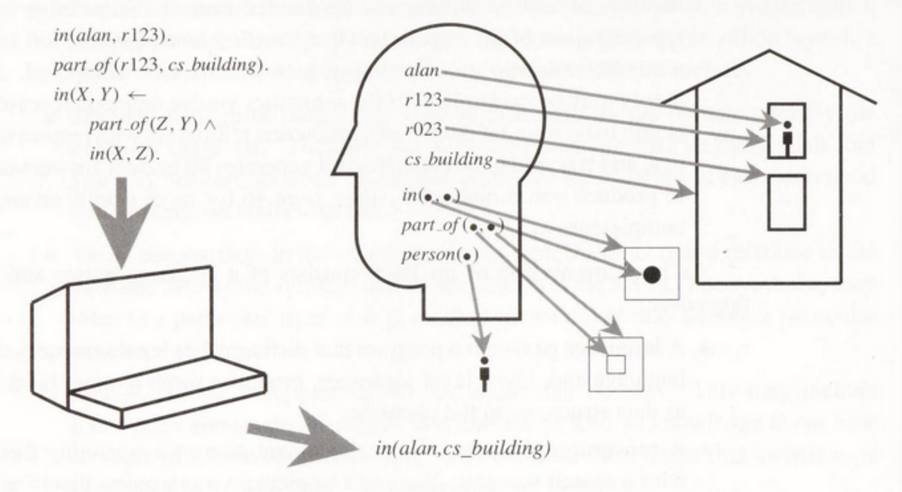
Knowledge Representation

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

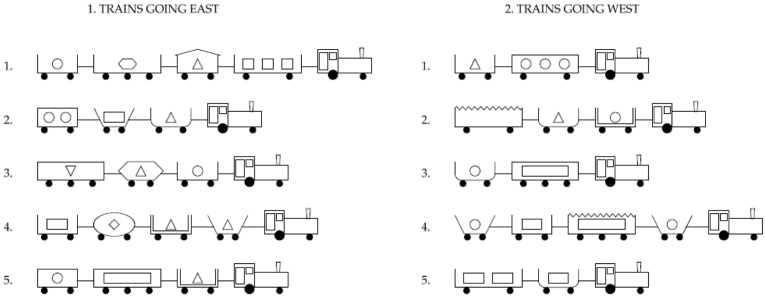
Why Knowledge Representation?

- To communicate and express “better” knowledge
- Uniform and standard notations
- Machine readable
- Reduce ambiguity

Knowledge Representation



Example: Michalski's trains



Deterministic Knowledge

- **positive:** "Subjects that snore have obstructive sleep apnea (OSA)"
- **negative**: "Subjects that do not snore do not have OSA"
- **unknown**: "Subjects that snore can have or not have OSA"

Uncertain Knowledge

- **positive:** “Subjects that snore have 70% of probability of having OSA”
- **negative**: “Subjects that do not snore have 70% of probability of not having OSA”
- **unknown**: “Portuguese subjects have 3% of probability of having OSA” (prevalence)

Representation

- Knowledge x Data?
- **Knowledge:** "symbolic representation of aspects of a universe of discourse"

Examples of "knowledge"

- John is a worker at UP
- All UP workers have salaries higher than 25,000 euros (:-)
- All UP workers know they must have a good lifestyle
- John does not think he has a good lifestyle
- All who know he should have a good lifestyle, but know he thinks he does not have a good lifestyle, are disappointed

Representation

- **Data:** symbolic representation of **simple** aspects of a universe of discourse
- **Data**: special case of "knowledge"

Examples of "data"

- John is married to Mary
- John is a UP worker
- The average salary at UP is 25,000 euros

Representation

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

Different formalisms

Natural language	Rules
Databases	Decision Trees
Frames	Logic
Scripts	Ontologies
Semantic networks	Causal networks
Genetic algorithms	Neural networks
Constraints	Markov models
Programming languages	Object oriented . . .

Representation in natural language

Texto Clínico

"Enviada por densidade assimétrica no QSE da mama esquerda. Esta alteração existe desde 2005 mas a avaliação ecográfica do exterior sugere a necessidade de biópsia. Exame mamário com alteração palpável com cerca de 30 mm no QSE da mama esquerda."

Disadvantages:

ambiguous, redundant, not structured, syntax and semantics not well understood. Dependent on context

Representation in databases

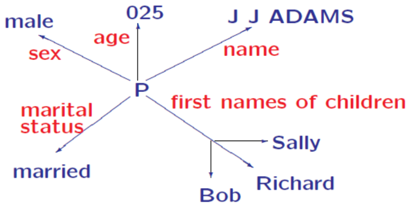
databases

```
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



Notes:

- only simple aspects can be represented (data)
- entity-relation model
- Reasoning = lookup

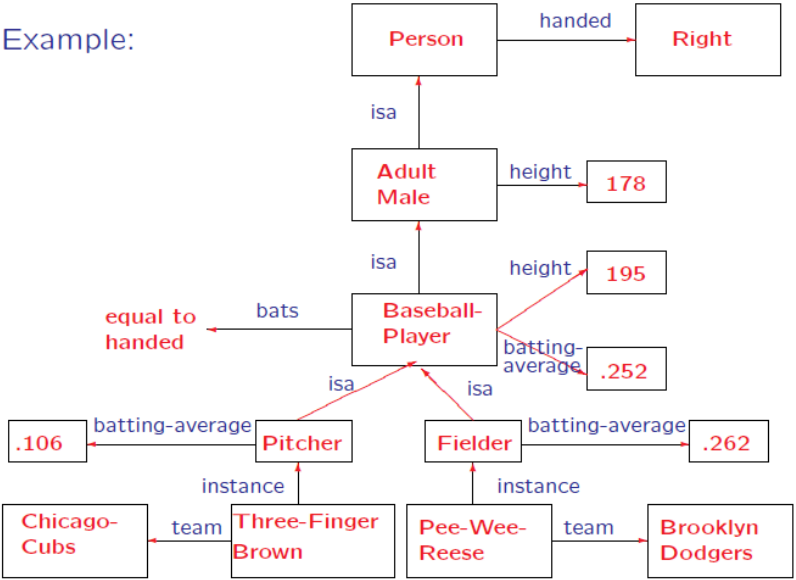
Representation in a single table

Usual: aggregate all tables in only one!

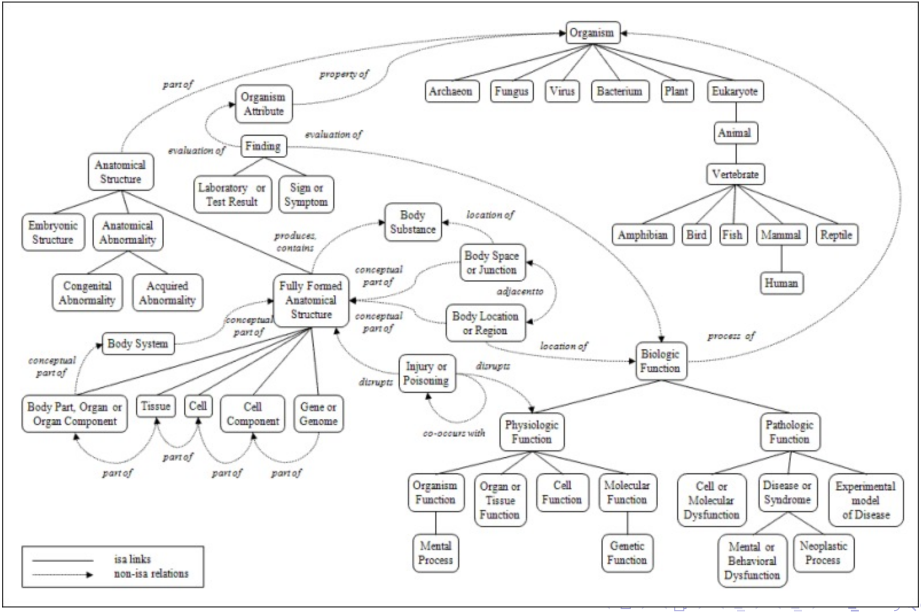
Patient	Location	Size	Date	Calcifications
P1	C	0.1	20050403	F, A
P1	C	0.2	20060412	F
P1	9	0.1	20060412	A
P2	12	0.3	20050415	M
...	...	...	...	...

Representation in semantic networks

Example:



Representation in semantic networks



Properties of semantic networks

- allows to structure knowledge
- assumes “default” values
- clear syntax, but semantics needs to be improved

Example based in Cyc (language)

"Bill Clinton belongs to the collection of U.S. presidents"

```
(#$isa #$BillClinton #$UnitedStatesPresident)
```

"All trees are plants"

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

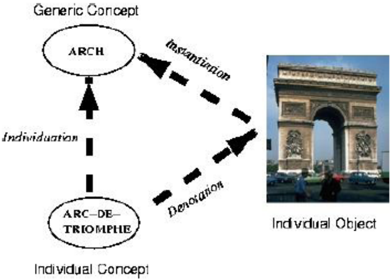
"Paris is the capital of France."

```
(#$capitalCity  #$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".

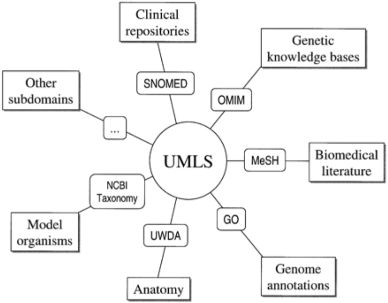
```
(#$implies
   (#$and
     (#$isa ?OBJ ?SUBSET)
     (#$genls ?SUBSET ?SUPERSET))
   (#$isa ?OBJ ?SUPERSET))
```

Important!



- distinguish:
  - concepts (representations) and objects (instances)
  - individual concepts and generalized concepts (schema)

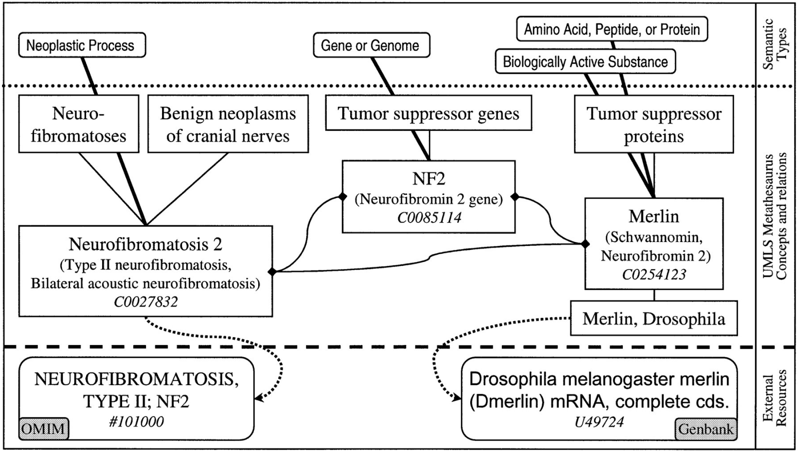
Unified Medical Language System – UMLS



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



[The Unified Medical Language System (UMLS): integrating biomedical terminology]()

XML

```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>
```

Representation procedural or declarative

- How? Procedural
- What? Declarative

Properties of good representations

- important/relevant objects and their relations are explicit
- express natural constraints
- represent objects and relations together
- omit irrelevant details
- transparent: easy to understand
- complete
- concise
- efficient to store and recover
- “computable”

Properties of good representations

- lexical part: determines which symbols must be used
- structural part: describes the format (constraints), how the symbols can be organized and put together
- procedural part: specifies access procedures that allow for the creation and modification of descriptions as well as pose queries
- semantics part: establish a way of associating a meaning to the description

Properties of good representations

for example, semantic networks:

- lexical: nodes, links, link labels
- structural: directed graph, with labeled arcs
- procedural: constructors, readers, writers, erasers (to create and modify the graph)
- semantics: meaning of nodes and arcs, depend on the 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

Logic representation

- Languages:
  - **syntax**: describes possible configs of the language that constitute valid sentences in the language
  - **semantic**: determines the meaning of a sentence
- example: $x > y$,
  - syntax: if x is a number and y is a number, then $x > y$ is a sentence over numbers
  - semantic: if $x > y$ returns true, else returns false

Logic representation

- Relations represented as tuples:

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

- Relations can depend on other relations:

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

- Symbols: and, if-then

Logic Representation

- Representations can be more complex to represent than simple queries to a database:

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

First order logic

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

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

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:

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

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

Introduction to logic

- Sentences are phrases that can be evaluated as true or false and can be used to form new sentences
- In propositional logic, grammar parts as:
  "it is not the case that..." **"and simultaneously"**
  are "connectives"
- atomic sentences or composed sentences
- variables: p, q, r, s ...

Introduction to logic

- connectives:
  - not ($\neg$ or $\sim$)
  - and ($\wedge$ or * or .)
  - or ($\vee$ or $+$),
  - implication ($\rightarrow$)
  - equivalence ($\leftrightarrow$)

Introduction to logic

connective	name	meaning
$\neg$	negation	it is not the case that
$\wedge$	conjunction	and
$\vee$	disjunction	or (inclusive)
$\rightarrow$	implication	if...then
$\leftrightarrow$	equivalence	if and only if

Introduction to logic: truth tables

AND

A	B	A $\wedge$ B
false	false	false
false	true	false
true	false	false
true	true	true

OR

A	B	A ∨ B
false	false	false
false	true	true
true	false	true
true	true	true

Introduction to logic: truth tables

NEGATION

A	¬ A
false	true
true	false

EQUIVALENCE

A	B	A $\leftrightarrow$ B
false	false	true
false	true	false
true	false	false
true	true	true

Lisboa é a capital de Portugal se e somente se 2 é par (!!)

IMPLICATION

- The value of an implication depends on the existence of some cause-effect relation between the antecedent and the consequent
- Paulo is ill. The doctor prescribes a drug to Paulo.
  If Paulo is ill, then the doctor prescribes a drug to Paulo.
  If the doctor prescribes a drug to Paulo then Paulo is ill. (?!)

IMPLICATION

- Usually, “if...then” does not form sentences whose value, in a given context, depends exclusively on the value of its components, in this context.
- Need to define a semantic to fix this flaw.

IMPLICATION

- In a mathematical notation, if...then is used when we want to present a condition that makes something true
- For example:
  A is a subset of B if and only if for all $u \in U$, if $u \in A$, then $u \in B$

IMPLICATION

U = {1,2,3}, A = {1} and B = {1,2}

It is clear that $A \subseteq B$

Then:

$$1 \in A \to 1 \in B$$

$$2 \in A \to 2 \in B$$

$$3 \in A \to 3 \in B$$

are all TRUE.

IMPLICATION

But $2 \in A$ and $3 \in A$ are false

Therefore:

$$TRUE \rightarrow TRUE : TRUE$$

$$FALSE \rightarrow TRUE : TRUE$$

$$FALSE \rightarrow FALSE : TRUE$$

Introduction to logic: truth tables

IMPLICATION

A	B	A $\rightarrow$ B
false	false	true
false	true	true
true	false	???
true	true	true

Is is common to consider *true* $\rightarrow$ *false* : *false*

Exercise: classify the following sentences as TRUE or FALSE

- Today is not Tuesday.
- Today is Friday and yesterday was Thursday.
- Today is Friday or tomorrow will not be Tuesday.
- If today is not Monday then tomorrow will not be Tuesday either.
- Today is not Friday if and only if yesterday was Thursday.

Introduction to logic: truth tables

Exercise

- First step: Determine the **atomic** sentences
  - Today is Tuesday
  - Today is Friday
  - Yesterday was Thursday
  - Tomorrow will be Tuesday
  - Today is Monday
- Second step: give them shorter names (not always good...): p, q, r, s, t

Exercise

- Third step: add connectives where needed and use parentheses
  - $(\neg p)$
  - $(q \wedge r)$
  - $(q \vee (\neg s))$
  - $((\neg t) \rightarrow (\neg s))$
  - $((\neg q) \leftrightarrow r)$

Proofs using truth tables

p	$\neg p$
false	true
true	false

Proofs using truth tables

q	r	$q \wedge r$
false	false	false
false	true	false
true	false	false
true	true	true

Proofs using truth tables

q	s	$\neg s$	$q \vee \neg s$
false	false	true	true
false	true	false	false
true	false	true	true
true	true	false	true

Proofs using truth tables

t	$\neg t$	s	$\neg s$	$((\neg t) \rightarrow (\neg s))$
false	true	false	true	true
false	true	true	false	false
true	false	false	true	true
true	false	true	false	true

Proofs using truth tables

q	$\neg q$	r	$((\neg q) \leftrightarrow r)$
false	true	false	false
false	true	true	true
true	false	false	true
true	false	true	false

Proofs using truth tables

Assuming that the sentences were said in a Friday:

$$\begin{array}{lcl} \neg p & : & \text{TRUE} \\ q \wedge r & : & \text{TRUE} \\ q \vee \neg s & : & \text{TRUE} \\ \neg t \rightarrow \neg s & : & \text{TRUE} \\ \neg q \leftrightarrow r & : & \text{FALSE} \end{array}$$

AND example

SNS24 sentence:

> *É obrigatório o uso de máscaras por todas as pessoas que circulem ou permaneçam nos espaços e vias públicas E sempre que o distanciamento físico recomendado pelas autoridades de saúde NÃO possa ser garantido.*

$$A \wedge \neg B \rightarrow C$$

A: circulem ou permaneçam...
B: distanciamento garantido
C: é obrigatório o uso de máscara

Exercise $\#1$

Represent the following sentences using propositional or first order logic:

- Example 1
  - All viral diseases are infectious diseases
  - AIDS 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