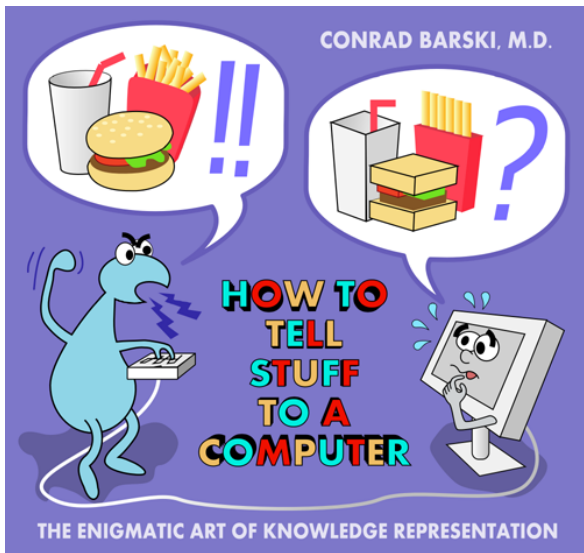


Knowledge Representation



Knowledge Representation: WHAT?

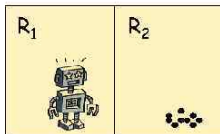
- To express knowledge that can be handled by a computer
 - ▶ Symbolic (logic-based, semantic networks, knowledge graphs, ontologies, etc)
 - ▶ Numerical (attribute-value, matrices, markov models, gaussian models etc)

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

State Representation



$$\text{In}(\text{Robot}, R_1) \wedge \text{Clean}(R_1)$$

Propositions
that "hold"
(i.e. are true)
in the state

Logical "and"
connective

6

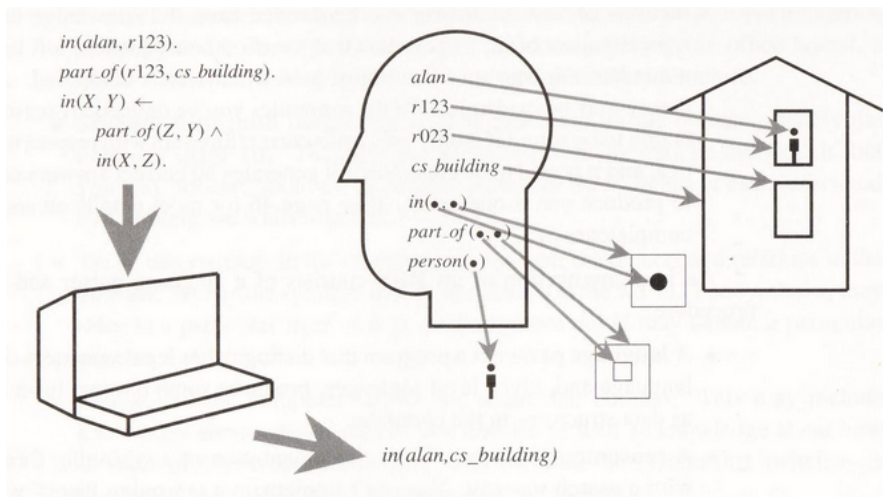
<https://present5.com/action-planning-where-logic-based-representation-of-knowledge-makes/>

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)

Representação

- **Conhecimento** x **Dados**?
- **Conhecimento**: “representação simbólica de aspectos de algum universo de discurso”

Exemplos de “conhecimento”

- José é um funcionário da UP
- Todos os funcionários da UP têm salários maiores que 25.000 euros (:-)
- Todos os funcionários da UP sabem que devem ter um bom estilo de vida
- José não acha que tem um bom estilo de vida
- Todos que sabem que José não acha que tem um bom estilo de vida estão desapontados

Representação

- **Dados:** “representação simbólica de aspectos **simples** de algum universo de discurso”
- **Dado:** caso especial de “conhecimento”

Exemplos de “dados”

- José é casado com Maria
- José é funcionário da UP
- O salário médio da UP é de 25.000 euros

Representação

- Representação do Conhecimento: expressar **conhecimento** de forma tratável pelo computador.

Diferentes formalismos

Linguagem natural	Regras
Bases de dados	Árvores de decisão
Frames	Lógica
Scripts	Ontologias
Redes Semânticas	Redes “causais”
Algoritmos genéticos	Redes neuronais
Restrições	Markov models
Linguagens	Orientação a objetos ...

Representação em *Linguagem Natural*

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.”

Desvantagens:

ambígua, redundante, pouca estrutura, sintaxe e semântica não são bem entendidas.

Representação em *Bases de Dados*

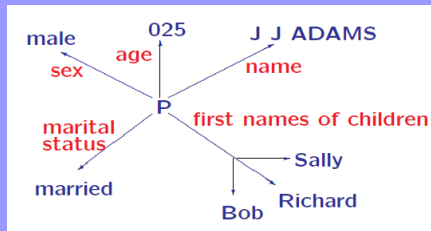
Base de Dados

```
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
           }
```

Representação em Bases de Dados: uma instância

Instância

J. J. ADAMS
025
male
married
Sally
Richard
Bob



Discussão

- apenas aspectos simples podem ser representados (dados)
- entidades e relações
- **Reasoning = lookup**



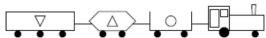


Representação em *uma única tabela*

Usual: agregar dados em uma única tabela!






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
...

Exercício

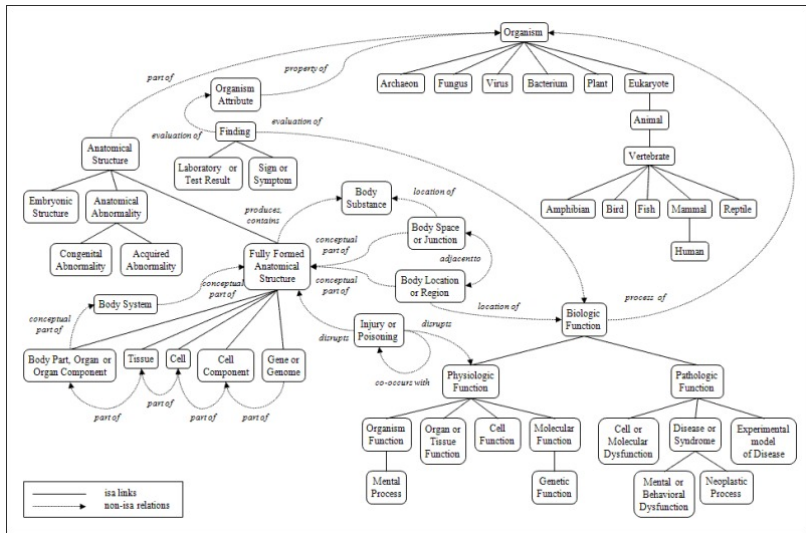
1. TRAINS GOING EAST

1. 
2. 
3. 
4. 
5. 

2. TRAINS GOING WEST

1. 
2. 
3. 
4. 
5. 

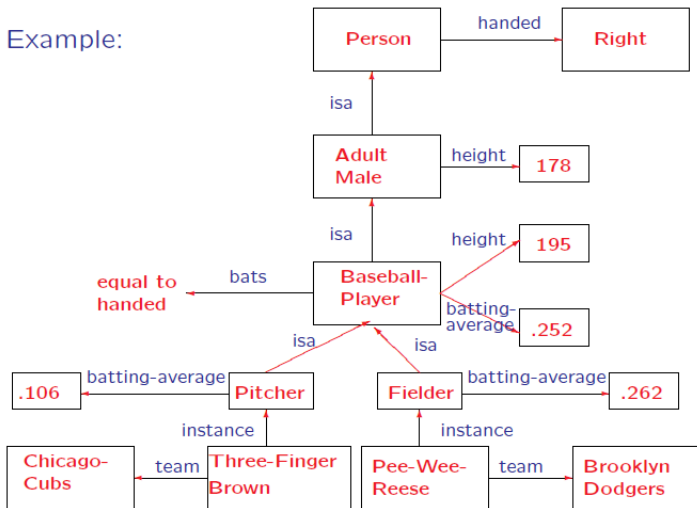
Representação em *Redes Semânticas*



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

Representação em Redes Semânticas

Example:



Propriedades de Redes Semânticas

- permite **estruturar** o conhecimento para refletir a parte do universo que está sendo representada
- valores “default”
- sintaxe clara, mas semântica precisa ser trabalhada

Exemplo baseado na linguagem Cyc

"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))
```

Frames

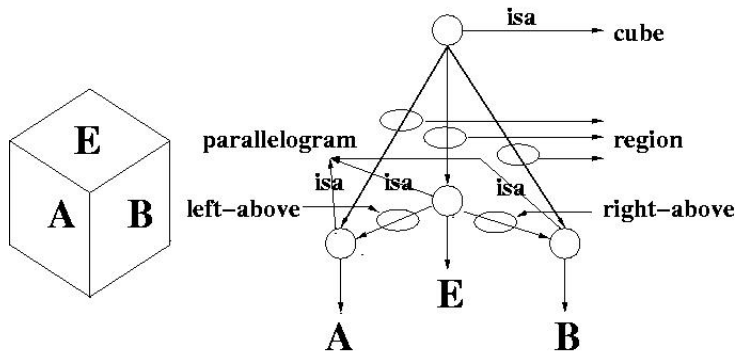
- Um “frame” consiste numa coleção de “slots”, cujo conteúdo pode ser um valor ou um apontador para outro Frame.

Festa de aniversário	
Vestuário:	social desportivo
Presente:	deve agradar o aniversariante deve ser comprado e embrulhado
Jogos:	escondidas colocar o rabo no burro
Decoração:	balões, brindes, papel crepe
Menu:	Bolo, Gelado, Refri, Cachorro quente
Bolo:	acender velinhas, assoprar velinhas, fazer um pedido, cantar parabéns
Gelado:	napolitano

Frames

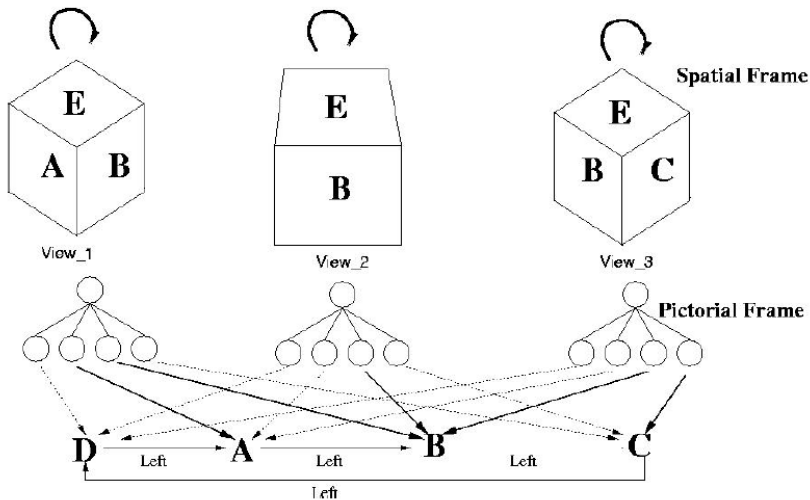
- Um “frame” consiste numa coleção de questões a serem respondidas sobre uma situação hipotética: especifica as questões e os métodos.
 - ▶ O que causou (agente)?
 - ▶ Qual é o propósito (intenção)?
 - ▶ Quais são as consequências (efeitos)?
 - ▶ A quem afeta (receptor)?
 - ▶ Como é feito (instrumentos/métodos)?

Exemplo



Objeto **composto** por relações.

Exemplo

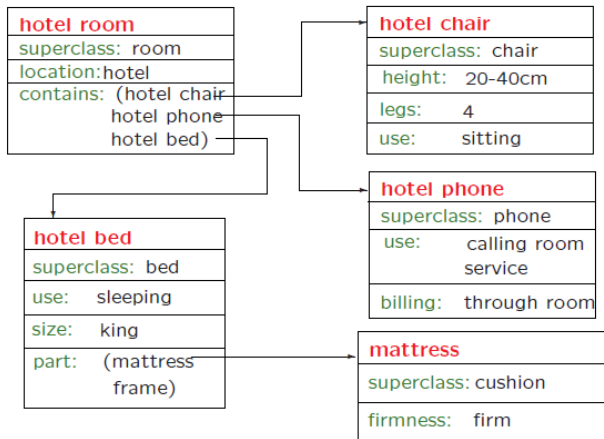


Diferentes aspectos de um cubo.

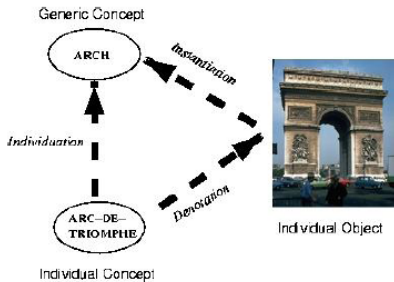
Frame para um aspecto do cubo

Um aspecto de um cubo usando representação em Frame

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)

Mais um exemplo

Importante!



- distinguir:
 - ▶ **conceitos** (representações) e **objetos** (instâncias)
 - ▶ **conceitos individuais** e **conceitos gerais**

Associação de procedimentos à representação

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

- Um “script” é uma representação estruturada que descreve uma sequência de eventos em um determinado contexto.
 - ▶ Estende os “frames” através de representação explícita de ações e mudanças de estados.
 - ▶ Define primitivas para descrever o universo:
 - PTRANS transferência física de um objeto (“go”)
 - ATRANS transferência de relações (“give”)
 - MTRANS transferência mental (“tell”)

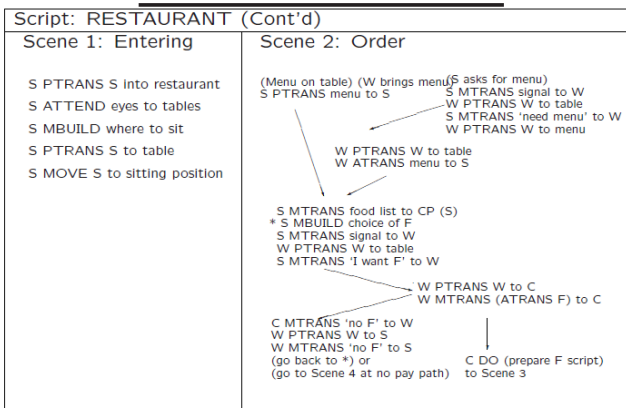
Exemplo de Script

Script para um restaurante

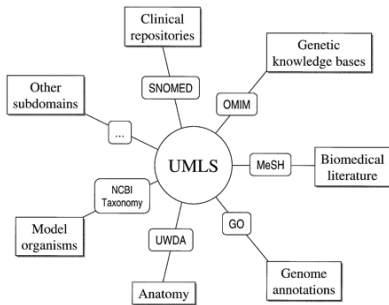
Script: RESTAURANT

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 para um restaurante (cont.)



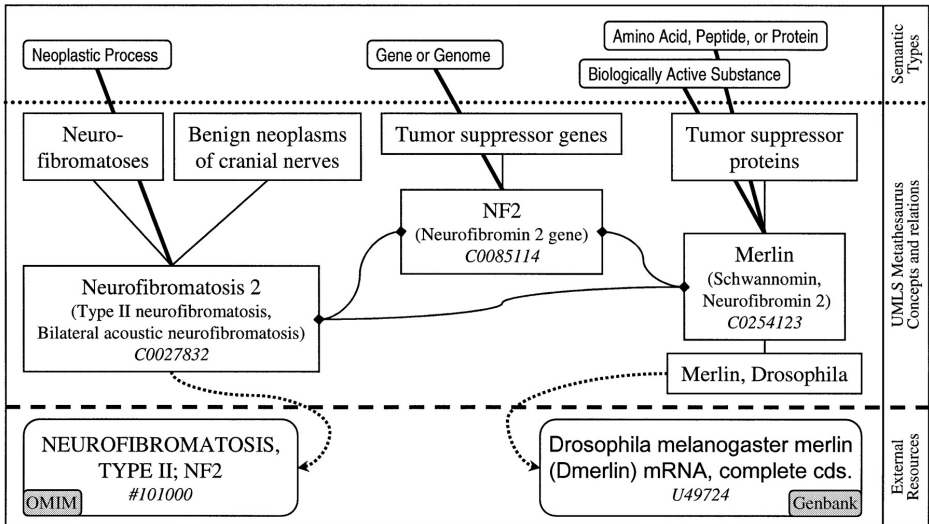
Unified Medical Language System – UMLS



MetaThesaurus

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



XML

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <Patient id="122">
3   <General>
4     <Weight-kg>35</Weight-kg>
5     <Height-cm>128</Height-cm>
6     <Day>05</Day>
7     <Month>02</Month>
8     <Year>2004</Year>
9     <Sex>1</Sex>
10    <AuscultationPosition>2</AuscultationPosition>
11  </General>
12  <SystemicPressure>
13    <SystemicPressureMethod>1</SystemicPressureMethod>
14    <SystolicSystemicPressure-mmHg>130</SystolicSystemicPressure-mmHg>
15    <DiastolicSystemicPressure-mmHg>90</DiastolicSystemicPressure-mmHg>
16  </SystemicPressure>
17  <PulmonaryPressure>
18    <PulmonaryPressureMethod />
19    <SystolicPulmonaryPressure-mmHg />
20    <DiastolicPulmonaryPressure-mmHg />
21    <CatheterismSimultaneousMeasurement />
22    <EchocardiograSameConsultation />
23  </PulmonaryPressure>
24  <Murmur>
25    <Cycle>2</Cycle>
26  </Murmur>
27  <S1>
28    <S1Status>1</S1Status>
29  </S1>
30  <S2>
31    <S2Status>1</S2Status>
32    <IfAbnormal>0</IfAbnormal>
33    <PulmonaryComponent>1</PulmonaryComponent>

```

Representação procedural ou declarativa

- Como? Procedural
- O que? Declarativo

Propriedades das boas Representações

- objetos importantes e suas relações estão explícitos
- expressam restrições que são naturais
- representam objetos e relações de forma conjunta
- omite detalhes irrelevantes
- transparente: fácil de entender
- completa
- concisa
- de armazenamento e recuperação rápidos
- “computáveis”

Propriedades das boas Representações

- Parte **léxica** que determina quais símbolos devem ser utilizados
- Parte **estrutural** que descreve a forma (restrições) como os símbolos podem ser organizados
- Parte **procedural** que especifica procedimentos de acesso que permitem a criação e modificação de descrições além de permitir fazer perguntas
- Parte **semântica** que estabelece uma forma de associar “significado” à descrição

Propriedades das boas Representações

Por exemplo, redes semânticas:

- Parte **léxica**: nodes, links, link labels
- Parte **estrutural**: grafo dirigido, com arestas etiquetadas
- Parte **procedural**: constructors, readers, writers, erasers (para criar e modificar o grafo)
- Parte **semântica**: significado dos nós e arestas depende da aplicação

Regras proposicionais

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
```

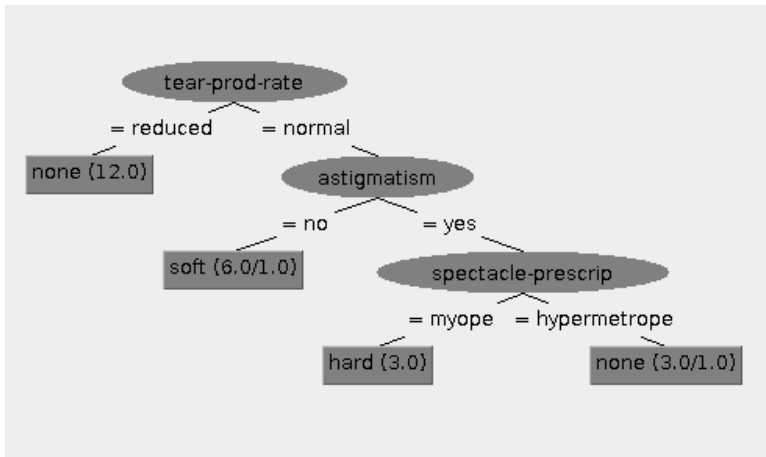
```
if: Hib.prior = 1 and not Hib_inactive and Hib1_age_in_months < 12
    and not Hib2_parameters_met
then: next.Hib2
```

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

Árvores de Decisão

- Simples e fácil de implementar.
- Recebe como entrada um objeto ou situação descrita por um conj de propriedades e produz uma resposta “sim” ou “não”. Representam funções booleanas.

Árvores de Decisão: Exemplo



Representação Lógica

- Linguagens:
 - ▶ **sintaxe**: descreve as possíveis configurações da linguagem que constituem sentenças válidas.
 - ▶ **semântica**: determina o significado de cada sentença.
- exemplo: $x > y$,
 - ▶ sintaxe: se x é um número e y é um número, então $x > y$ é uma sentença sobre números.
 - ▶ semântica: se $x > y$ retorna verdadeiro, senão retorna falso.

Representação Lógica

- Relações representadas como tuplos:

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

- Relações podem depender de outras relações:

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

- Símbolos: and, or, if-then

Representação Lógica

- Representações podem ser mais complexas de representar:
IF lesion(x) AND connected(y,x) THEN lesion(y)

Regras de primeira ordem

$$\begin{aligned}
 \text{same_finding}(F_1, F_2) \leftarrow & \text{MLOView}(F_1) \wedge \text{CCView}(F_2) \wedge \\
 & \text{nipple_distance}(F_1, D_1) \wedge \text{nipple_distance}(F_2, D_2) \wedge \\
 & (\text{abs}(D_1 - D_2) < \epsilon) \wedge \\
 & \text{side}(F_1, \text{left}) \wedge \text{side}(F_2, \text{left}) \wedge \\
 & \text{quadrant}(F_1, \text{upper_outer}) \wedge \text{quadrant}(F_2, \text{upper_outer}) \wedge \\
 & \text{massShape}(F_1, \text{oval}) \wedge \text{massShape}(F_2, \text{oval}).
 \end{aligned}$$

$$\begin{aligned}
 \text{previous_finding}(F_1, F_2) \leftarrow & \text{mammo}(P, F_1) \wedge \text{mammo}(P, F_2) \wedge \\
 & \text{date}(F_1, D_1) \wedge \text{date}(F_2, D_2) \wedge \\
 & (D_1 < D_2 \vee D_2 < D_1)
 \end{aligned}$$

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:

$$\begin{aligned}
 \text{is_malignant}(A) \leftarrow & \text{mass}(A, \text{present}) \wedge \text{previous_finding}(A, B) \wedge \\
 & (\text{massSize}(A) < \text{massSize}(B)) \wedge \text{calc}(B, \text{present}) \wedge \\
 & \text{previous_finding}(A, C) \wedge \text{calcFineLinear}(C, \text{yes})
 \end{aligned}$$

source: *Automated Diagnosis of Breast Cancer on Medical Images, in Foundations of Biomedical*

Exercício 1

Represente as seguintes frases usando lógica de primeira ordem:

- 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

Exercício 2

Representar as frases abaixo em lógica de primeira ordem:

- 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 é pai de Silver.
- g) descendente e pais são relações inversas.
- h) Todo mamífero tem pais.