

**Ústav lékařské informatiky, 2. LF UK**

# **Formalizace znalostí a znalostní systémy**

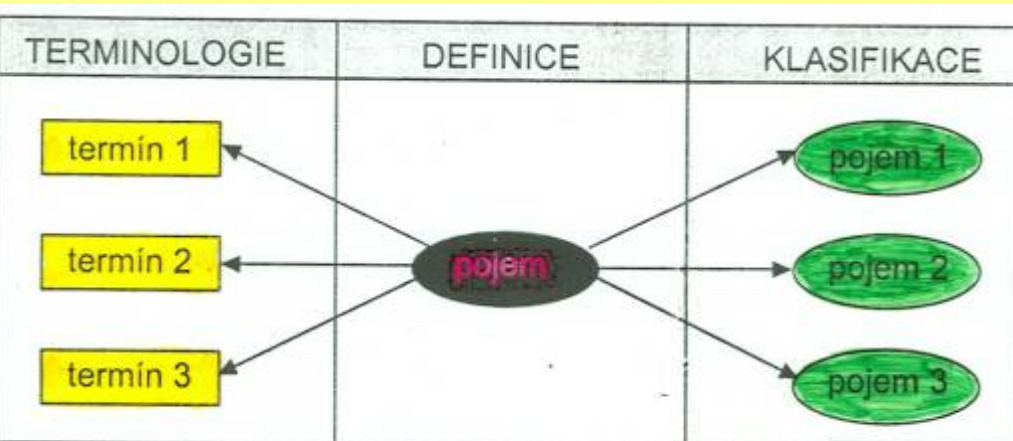
**2005**



# Lékařské pojmy

Jazyk přirozený, formální a odborný

Nomenklatura – utříděný systém pojmů



Formální popis pojmu

Lékařská terminologie  
- synonyma

Synonyma – pravidla pro výběr  
standardní ukládání do DB

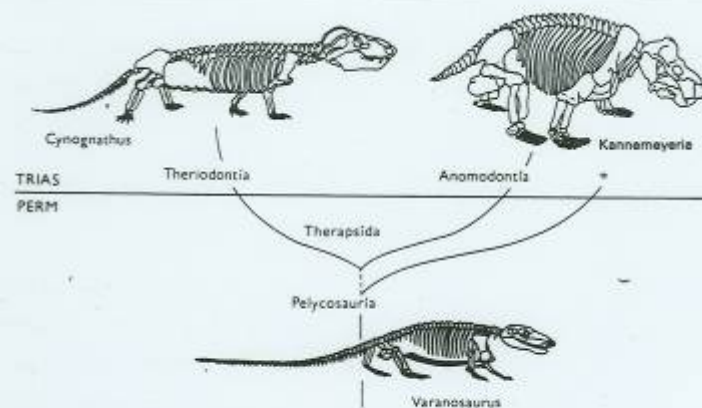
## D 45870 Anafylaktoidní purpura

- Purpura DEBRÉ-LAMY
- Esenciální atrombocytopenická purpura (SCHULTZE)
- Heberdenův-Willianův syndrom
- Kapilarotoxická purpura (FRANK)
- Peliosis rheumatica
- Purpura allergica
- Purpura anaphylactica (GLANZMANN)
- Morbus Schönlein-Henoch
- Imunní vaskulitis

# Požadavky na nomenklaturu

<b>SYSTÉMOVOST</b>	
Mezinárodní před-, přípona	LYMFO granulOM
Eponymum	Morbus Hodgkin
<b>PŘESNOST</b>	
Adherentní vlastnost	Anafylaktoidní purpura
Inherentní vlastnost	Imunní vaskulitis
<b>ÚSPORNOST</b>	
Složené slovo	Imunodefekt
Složený výraz	Imunodeficitní stav
<b>FUNKČNOST</b>	
Metaforie	Přirození zabíječi
Kódové značení	CD7 T- lymfocyty

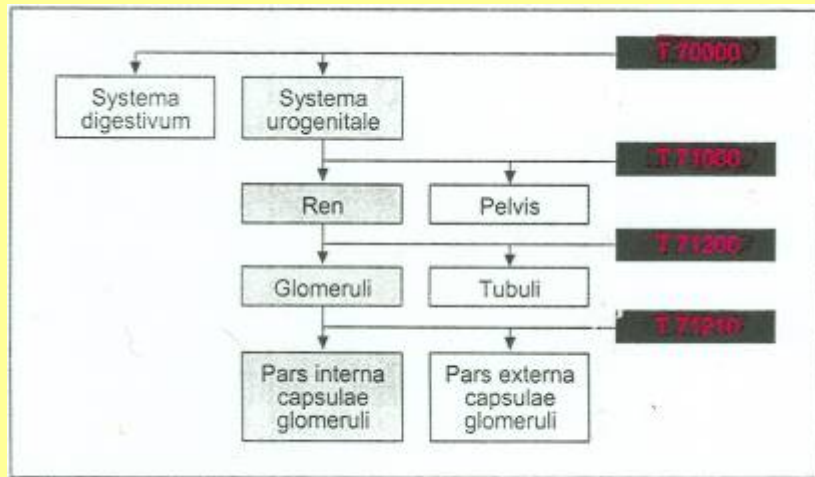
Mezinárodní pravidla  
zoologické nomenklatury



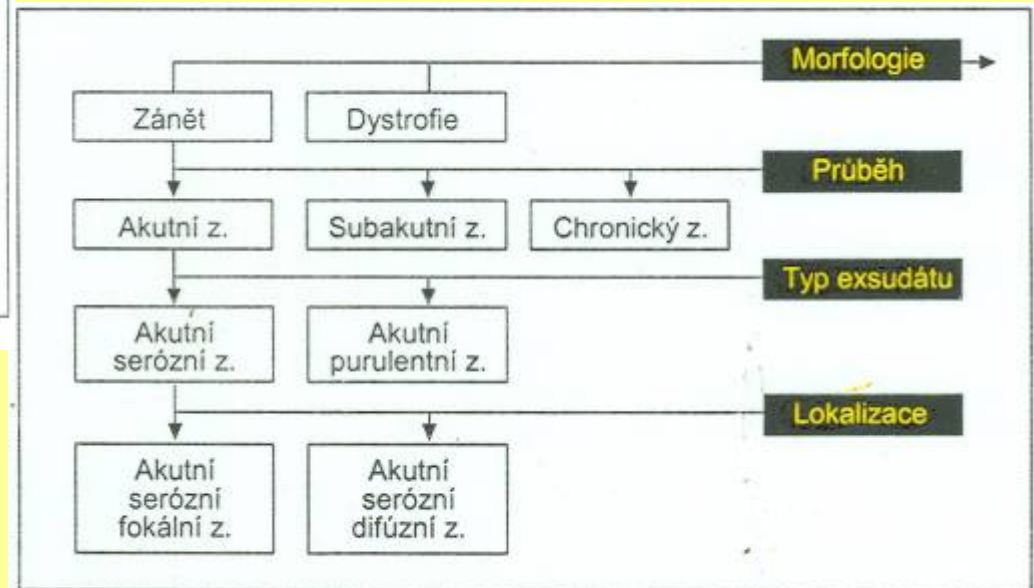
# Klasifikační systémy

**Hierarchie** – formalizace logických vztahů, které mají charakter souřadnosti – pořadnosti

- a) hierarchie partitivní (celek – část)
- b) hierarchie generická (abstrakce)

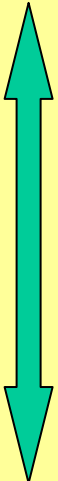


Partitivní hierarchie



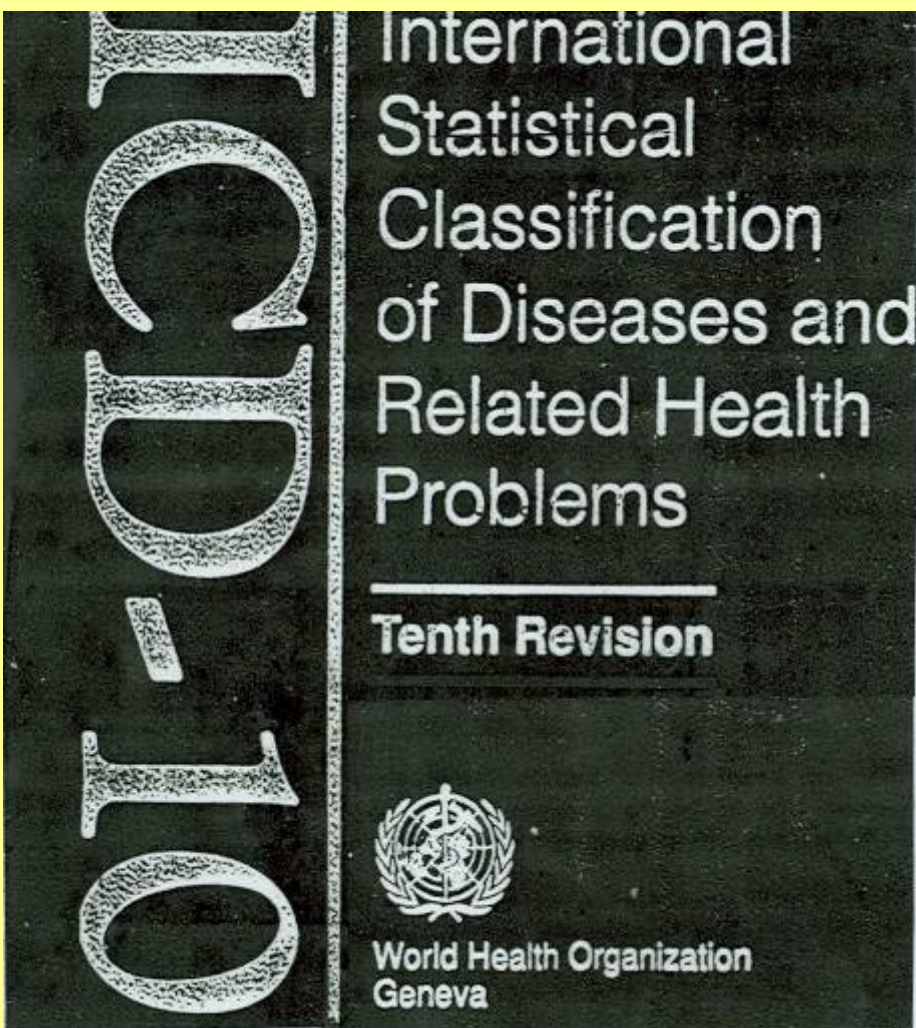
Generická hierarchie, na každé etáži vždy 1-dimenzionální

# Klasifikace - Rozdělení kódování systémů

• 1) Bibliografie	MeSH		UMLS
• 2) Diagnózy	ICD		
• 3) Procedury	ICPM, VZP		
• 4) Znalosti	SNOMED		



# Klasifikace ICD



## **B37** Kandidóza - candidiasis

- .0 Kandidová stomatitida
- .1 Plicní kandidóza
- .2 Kandidóza kůže a nehtů
- .3+ Kandidóza vulvy a pochvy (N77.1\*)
- .4+ Kandidóza jiných močových a pohlavních lokalizací
- .5+ Kandidová meningitida (G02.1\*)
- .6+ Kandidová endokarditida (I39.8\*)
- .7 Kandidová septikémie
- .8 Kandidóza jiných lokalizací
- .9 Kandidóza, NS

## **I39\*** Zánět srdeční nitroblány - endokarditis při nemocech zařazených jinde

*Patří sem: postižení endokardu při:*

- kandidové infekce (B37 . 6 +)
- gonokokové infekce (A54 . 8 +)
- revmatoidní artritidě (M05 . 3 +)

- .0 \* Nemoci dvojčipé chlopně
- .1 \* Onemocnění aortální chlopně
- .2 \* Onemocnění trojčipé chlopně
- .3 \* Onemocnění pulmonální chlopně
- .4 \* Mnohočetná onemocnění chlopní
- .5 \* Endokarditis neurčené chlopně

# Klasifikace – MeSH, SNOMED

GENITAL NEOPLASMS	C4.588.945	C12.878	C13.371.830
ENITAL NEOPLASMS, FEMALE	C4.588.945.418	C13.371.210	
FALLOPIAN TUBE NEOPLASMS	C4.588.945.418.365	C13.371.56	
OVARIAN NEOPLASMS	C4.588.945.418.685	C4.588.322	C13.371.56
KRUKENBERG'S TUMOR	C4.588.945.418.685.538	C4.587.117	C4.587.576
MEIGS' SYNDROME	C4.588.945.418.685.595	C13.371.56	C4.588.322
UTERINE NEOPLASMS	C4.588.945.418.948	C4.588.322	C13.371.56
CERVIX NEOPLASMS	C4.588.945.418.948.170	C13.371.852	
CERVIX DYSPLASIA	C4.588.945.418.948.170.170	C13.371.852	
VAGINAL NEOPLASMS	C4.588.945.418.955	C4.834.160	C13.371.852
VULVAR NEOPLASMS	C4.588.945.418.968	C13.371.894	
ENITAL NEOPLASMS, MALE	C4.588.945.418.968	C13.371.946	
PENILE NEOPLASMS	C4.588.945.440.715	C12.294.160	
PROSTATIC NEOPLASMS	C4.588.945.440.770	C12.294.494	
TESTICULAR NEOPLASMS	C4.588.945.440.915	C12.294.565	
ROLOGIC NEOPLASMS	C4.588.945.947	C4.588.322	C12.294.829
BLADDER NEOPLASMS	C4.588.945.947.125	C12.878.950	
KIDNEY NEOPLASMS	C4.588.945.947.535	C12.777.103	
URETERAL NEOPLASMS	C4.588.945.947.940	C12.777.419	
URETHRAL NEOPLASMS	C4.588.945.947.945	C12.777.725	
ENEREAL TUMORS, VETERINARY	C4.588.945.956	C12.777.767	
		C12.950	
KIDNEY FAILURE, CHRONIC	C12.777.419.453		
KIDNEY NEOPLASMS	C12.777.419.473	C4.588.945	
KIDNEY PAPILLARY NECROSIS	C12.777.419.493		
NEPHRITIS	C12.777.419.570		
GLOMERULONEPHRITIS	C12.777.419.570.363		
GLOMERULONEPHRITIS, IGA	C12.777.419.570.363.608	C10.111.525	
GLOMERULONEPHRITIS, MEMBRANOPROLIFERATIVE	C12.777.419.570.363.615	C10.425	
GLOMERULONEPHRITIS, MEMBRANOUS	C12.777.419.570.363.625	C10.111.525	
GLOMERULOSCLEROSIS, FOCAL	C12.777.419.570.363.640		
GOODPASTURE'S SYNDROME	C12.777.419.570.363.651	C8.381.314	C20.111.550
LUPUS NEPHRITIS	C12.777.419.570.363.680	C20.111.590	C23.205.200
NEPHRITIS, HEREDITARY	C12.777.419.570.620	C16.131.939	
NEPHRITIS, INTERSTITIAL	C12.777.419.570.643		
BALKAN NEPHROPATHY	C12.777.419.570.643.150		
PYELONEPHRITIS	C12.777.419.570.643.790	C12.777.419	
PYELONEPHRITIS, XANTHOGRANULOMATOUS	C12.777.419.570.643.790.810		
NEPHROCALCINOSIS	C12.777.419.590	C12.777.419	
NEPHROSCLEROSIS	C12.777.419.610	C18.452.174	
NEPHROSIS	C12.777.419.630	C16.907.489	

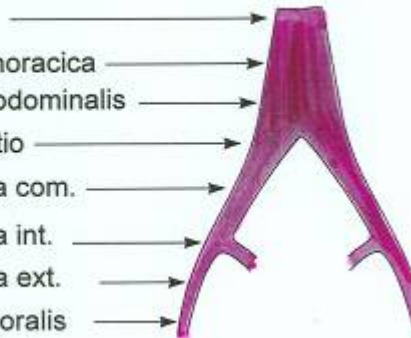
## MORFOLOGIE

M52100 Ateroskleróza	
M34000 Obstrukce	
M32400 Aneurysma	
M34100 Stenosa	
M35100 Thrombus	
M35300 Embolie	
M40000 Zánět	



## TOPOGRAFIE

42000 Aorta	
42400 Pars Thoracica	
42500 Pars Abdominalis	
42580 Bifurcatio	
46710 A. Iliaca com.	
46740 A. Iliaca int.	
46910 A. Iliaca ext.	
47400 A. Femoralis	





# Klasifikace – SNOMED, UMLS

**TOPOGRAFIE** ledvina T71000

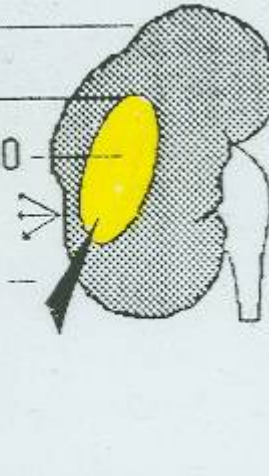
**MORFOLOGIE** absces M417100

**ETIOLOGIE** Staphylococcus E24400

**FUNKCE** bolest F82600

**PROCEDURA** nefrotomie P10000

**SYNDROM** sepsa S00800



**Prostatectomy** MC0033573

LT ☐ NH ☐ SP noun/noun phrase

TYP Therapeutic or Preventive Procedure

MSH/MH/D011468  
SNM/RT/P-1100  
MTH/RT/748  
LCH/PT

**CONCEPT DEFINITION**

[B] Excision, NOS (-ectomy)  
[N] Transurethral cryosurgical removal of prostate (postoperative irrigations and aspiration of sloughing tissue included)

**SOURCE CONTEXT for MSH/MH/D011468**

Analytical, Diagnostic and Therapeutic Techniques and Equipment (MeSH Category)

Surgery, Operative

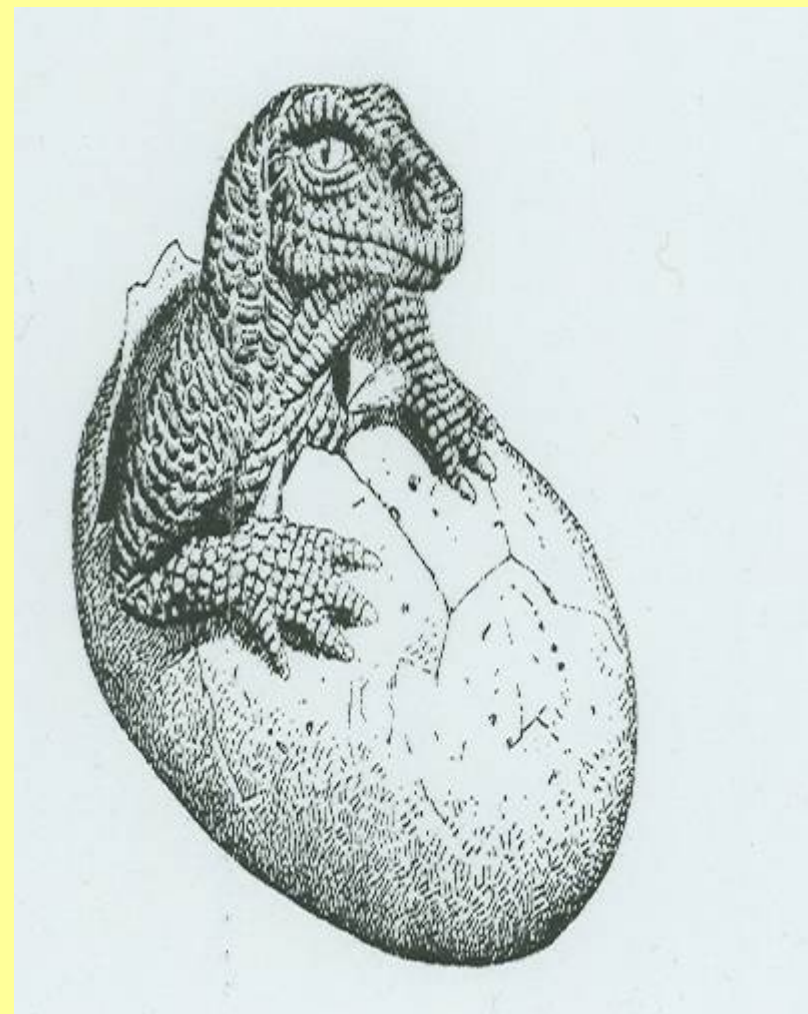
Surgery, Urogenital (Non MeSH)

Surgery, Urologic (Non MeSH)

Surgery, Urologic, Male (Non MeSH)

==> Prostatectomy

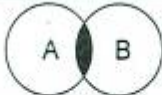





Terms Occurs MeSH Index LinkTo Note Locate

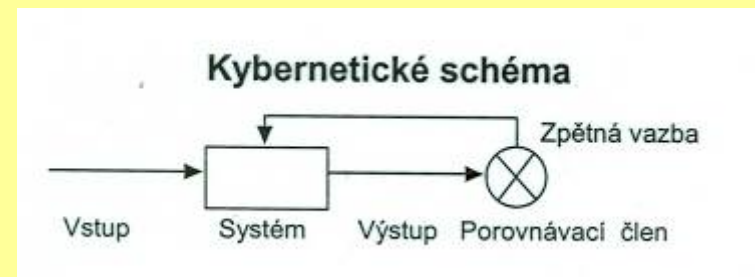




# Formalizace problému

	OBJEKT	VZTAH	NEJISTOTA
LOGIKA	výrok	predikát	pravdivost
TEORIE MNOŽIN	množina	relace	příslušnost
KYBERNETIKA	systém	informace	entropie

Množinová notace	Vennův diagram	Logická notace
$A \cap B$ průnik		$A \wedge B$ konjunkce „a“ „&“
$A \cup B$ sjednocení		$A \vee B$ disjunkce „nebo“
$B \subset A$ podmnožina $B \in A$ prvek		$A \Rightarrow B$ implikace „jestliže-potom“
$A \equiv B$ totožnost		$A \Leftrightarrow B$ ekvivalence „právě tehdy když“
$\bar{A}$ doplněk		$\neg A$ negace „ne“
$A \setminus B$ rozdíl		$A \wedge \neg B$ složený logický prvek



# Pravidla - Léčba

## PRAVIDLA - LÉČBA



Terapie



Stav (Příznak, Diagnóza)

Typ pravidla		Příklad
Dávkování	$T \Rightarrow d$ 	Phenobarbital $\Rightarrow$ 1-6 mg/kg/den
	$T1 \& T2 \Rightarrow d'$ 	Heparin & Salicyláty $\Rightarrow$ změna dávky
	$T \& S \Rightarrow d'$ 	Uremia & Gentamycin $\Rightarrow$ změna dávky
Interakce	$T1 \Rightarrow T2$ 	Tetracyclin $\Rightarrow$ Vitamin B
	$T1 \Rightarrow \neg T2$ 	Digoxin $\Rightarrow$ not Ephedrin
Indikace	$S \Rightarrow T$ 	Anaphylaxis $\Rightarrow$ Hydrocortizon
	$S \Rightarrow T1 \& T2$ 	Asthmatický záchvat $\Rightarrow$ Sympatomimetika or Aminophylin
Kontraindikace	$S \Rightarrow \neg T$ 	Glaucoma $\Rightarrow$ not Salbutamol
Vedlejší účinky	$T \Rightarrow_w S$ 	Phenylbutazon $\Rightarrow$ Agranulocytosis

# Pravidla - Diagnostika








## PRAVIDLA - DIAGNOSTIKA



Příznak



Diagnóza

Typ	pravidla	Příklad
Nutný příznak	$D \leftrightarrow S$ 	Lymphogranuloma $\leftrightarrow$ Sternbergovy buňky
	$S \Rightarrow D ; D \Rightarrow_w S$ 	Kaiserův Fleischerův prstenec $\Rightarrow$ Wilsonova nemoc
	$D \Rightarrow \neg S$ 	Brutonova agamaglobulinémie $\Rightarrow$ not defekt T-Lymf.
Postačující příznak	$D \Rightarrow S ; S \Rightarrow_w D$ 	Nefrotický syndrom $\Rightarrow$ proteinurie
	$D \Rightarrow S1 \vee S2$ 	3 Jonesova kriteria $\Rightarrow$ revmatická horečka
Přispívající příznak	$D \Rightarrow_w S ; S \Rightarrow_w D$ 	Ikterus $\Rightarrow$ Hepatitis
	$S \Rightarrow_w \neg D$ 	Pankreatitis $\Rightarrow$ not Hemoptoe

# Transformace pravidel – umělá inteligence



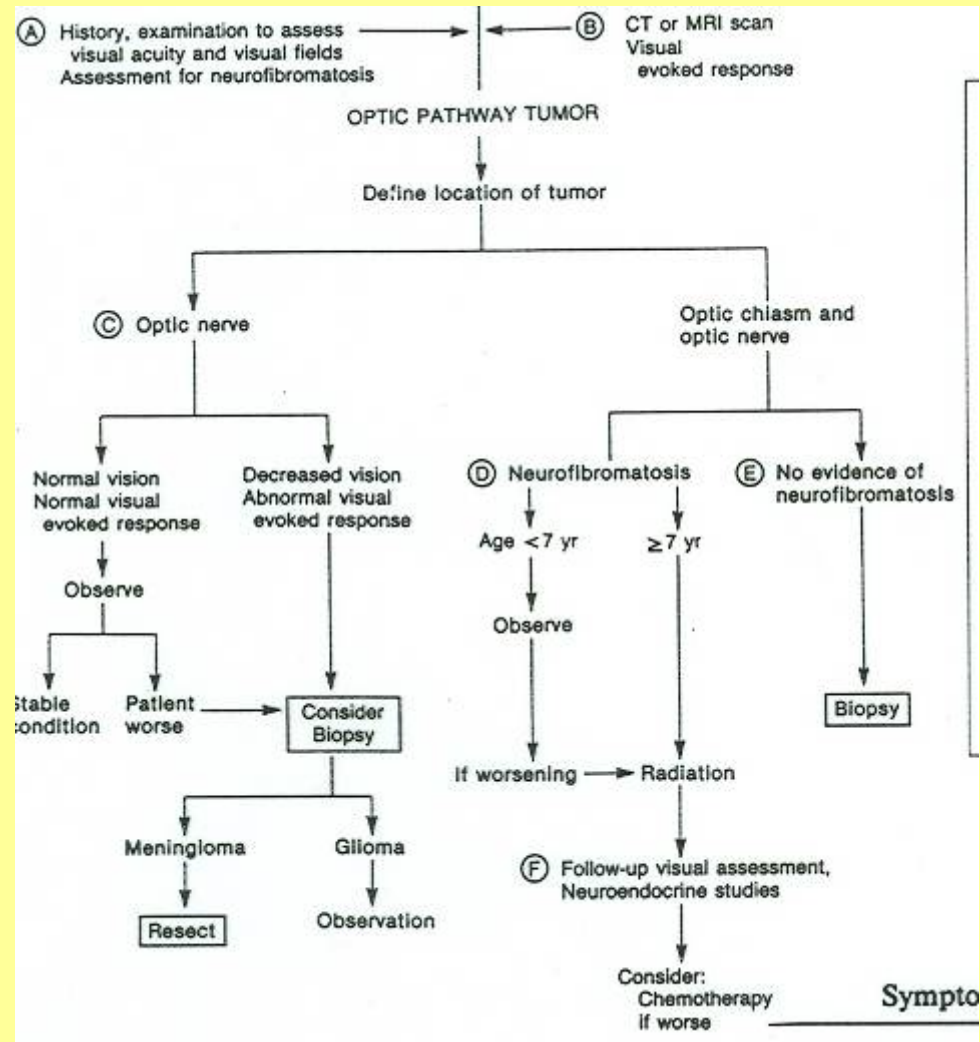


# Využití znalostních systémů v klinické praxi

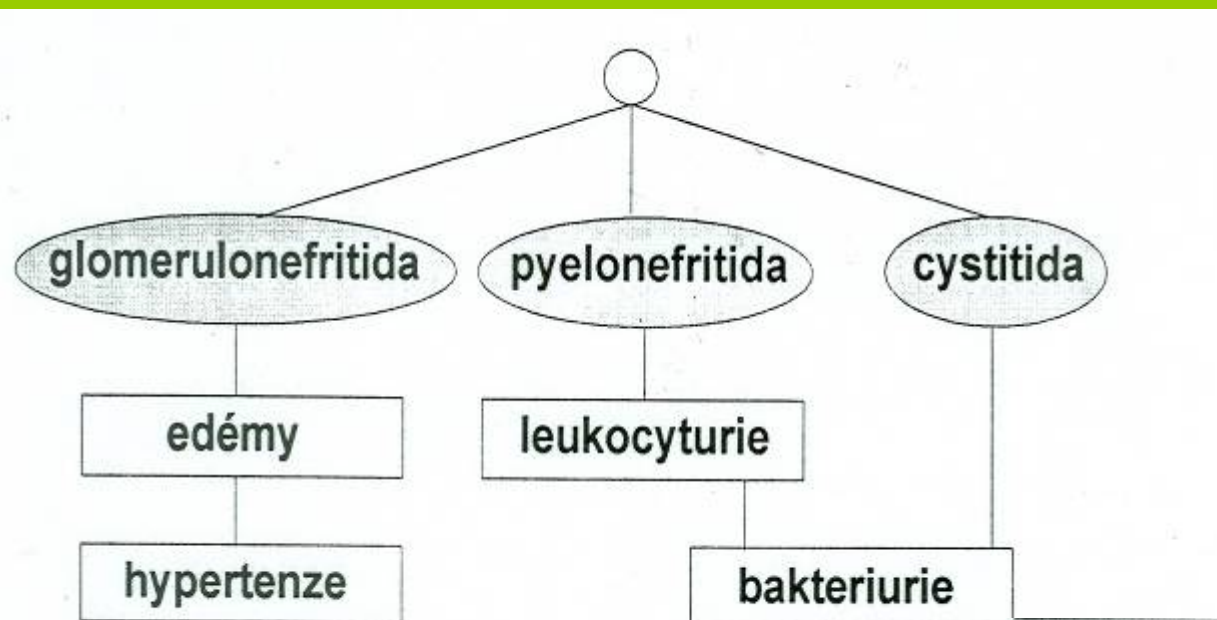
## Expertní systémy

Název	Využití
MYCIN (Stanford)	Léčba infekcí
MYOPAT (Praha)	Neurologie
QMR (Pittsburg)	Interna
ILLIAD (Salt Lake City)	Všeobecné lékařství

# Algoritmičtý model



# Logický model - Programování v jazyce Prolog



?	nález (glomerulonefritida, X)
	x = edémy x = hypertenze 2 solutions

Nález (Dg, Příznak)	
	nález (pyelonefritida, bakteriurie) nález (pyelonefritida, leukocyturie) nález (glomerulonefritida, edémy) nález (glomerulonefritida, hypertenze) nález (cystitida, bakteriurie)

?	nález (Dg, bakteriurie)
	Dg = pyelonefritida Dg = cystitida 2 solutions

# Pravděpodobnostní model

Code	Symptoms
BW	$x_1$ = age 1 mo. to 1 yr.
BW	$x_2$ = age 1 to 20 yr.
BW	$x_3$ = > 20 yrs.
BW	$x_4$ = cyanosis, mild
BW	$x_5$ = cyanosis, severe (with clubbing)
BW	$x_6$ = cyanosis, intermittent
BW	$x_7$ = cyanosis, differential
BW	$x_8$ = squatting
BW	$x_9$ = dyspnea
BW	$x_{10}$ = easy fatigue
BW	$x_{11}$ = orthopnea

$y_1$  = normal  
 $y_2$  = atrial septal defect without pulmonary stenosis or pulmonary hypertension\*  
 $y_3$  = atrial septal defect with pulmonary stenosis  
 $y_4$  = atrial septal defect with pulmonary hypertension\*  
 $y_5$  = complete endocardial cushion defect (A-V commune)  
 $y_6$  = partial anomalous pulmonary venous connections (without atrial septal defect)  
 $y_7$  = total anomalous pulmonary venous connections (supradiaphragmatic)  
 $y_8$  = tricuspid atresia without transposition  
 $y_9$  = Ebstein's anomaly of tricuspid valve  
 $y_{10}$  = ventricular septal defect with valvular pulmonary stenosis

Diseases	Incidence	Symptoms															
		$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$x_{13}$	$x_{14}$	$x_{15}$	$x_{16}$
$y_1$	0.100	01	49	50	01	00	01	00	01	01	10	03	05	05	03	05	01
$y_2$	.081	10	50	50	02	01	02	00	01	35	50	05	02	40	01	02	02
$y_3$	.005	30	60	10	20	10	20	00	01	60	70	05	02	10	10	02	02
$y_4$	.001	10	20	70	30	10	25	00	01	80	90	05	05	15	10	02	02
$y_5$	.027	20	50	30	15	05	10	00	01	40	50	05	05	30	05	60	15
$y_6$	.005	10	40	50	01	01	01	00	01	15	20	01	05	05	01	02	02
$y_7$	.001	20	70	10	65	10	05	00	01	70	80	05	05	20	05	02	02
$y_8$	.018	50	48	02	30	65	01	00	10	80	90	20	05	15	10	02	05
$y_9$	.001	10	45	45	22	44	01	00	22	80	80	10	30	15	22	05	25
$y_{10}$	.054	40	55	05	25	25	10	00	30	75	90	05	05	10	20	02	02
$y_{11}$	.063	40	55	05	30	30	10	00	40	75	90	05	05	10	25	02	02
$y_{12}$	0.45	20	70	10	01	01	01	00	01	50	65	01	01	01	10	02	02
$y_{13}$	.013	20	70	10	01	01	01	00	01	50	65	01	01	01	10	02	02

A Mathematical Approach  
to Medical Diagnosis:  
Application to Congenital  
Heart Disease

Homer R. Warner, Alan  
F. Toronto, L. George  
Veasey, and Robert  
Stephenson



# Skórovací model I.- skóre zanlostí

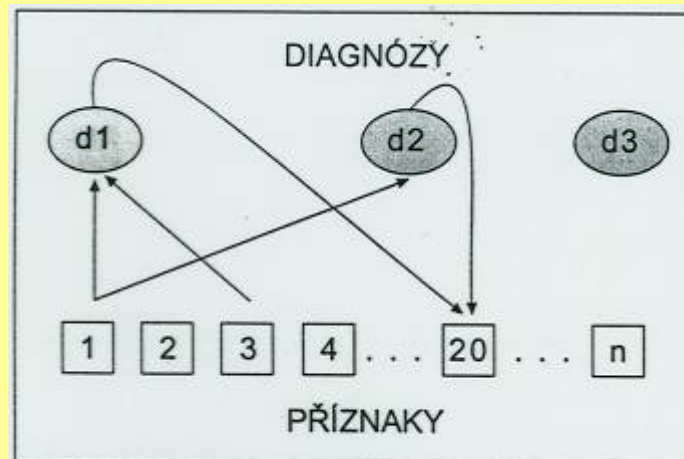
## INTERNIST - skórování

Specifita (jak často je Dg u sympt. přítomná)	Senzitivita (jak často je přítomen sympt. u Dg)
0 - extrémně vzácně	0 - nikdy
1 - vzácně	1 - vzácně
2 - často	2 - v menšině případů
3 - většinou	3 - v polovině případů
4 - v převaze případů	4 - ve většině případů
5 - prakticky vždy	5 - prakticky vždy

BOLESTI V PODBŘÍŠKU →<sub>2</sub> APPENDICITIS  
 APPENDICITIS →<sub>5</sub> BOLESTI V PODBŘÍŠKU

<b>DIAGNÓZA:</b> nefrotický syndrom	
<b>NUTNÝ PŘÍZNAK:</b> proteinurie	
<b>POSTACUJÍCÍ PŘÍZNAK:</b> edemy a proteinurie >5g/24hod	
<b>HLAVNÍ SKÓRE:</b>	
koncentrace albuminu v seru	<b>Váhy příznaků</b>
nizka:	1.0
vysoka:	-1.0
proteinurie	
>5g/24hod:	1.0
>1g <5g/24hod:	0.5
otoky	
symetrické a masivní:	1.0
symetrické ale ne masivní:	0.3
asymetrické:	-0.5
chybí:	-1.0
<b>POMOCNÉ SKÓRE:</b>	
koncentrace cholesterolu v seru	
vysoka:	1.0
ne vysoka:	-1.0
lipidy v moči	
pozitivní:	1.0
negativní:	-0.5

Úsek báze znalostí expertního systému



Funkce kognitivního systému

# Skórovací model II. - interakce

----- PATIENT - 1 -----

(1) Patient's name: (first-last)

**\*\* FRED SMITH**

*{background data}*

(2) Sex:

**\*\*MALE**

(3) Age:

**\*\*55**

(4) Have you been able to obtain positive cultures from a site at which Fred Smith has an infection?

**\*\*YES**

----- INFECTION - 1 -----

(5) What is the infection?

**\*\*PRIMARY-BACTEREMIA**

(6) Please give the date and approximate time when signs or symptoms of the primary-bacteremia (INFECTION-1) first appeared. (mo/da/year)

**\*\*MAY 5, 1995**

The most recent positive culture associated with the primary-bacteremia (INFECTION-1) will be referred to as:

----- CULTURE - 1 -----

(7) From what site was the specimen for CULTURE-1 taken?

**\*\*BLOOD**

(8) Please give the date and time when this blood culture (CULTURE-1) was obtained. (mo/da/year)

**\*\*MAY 9, 1995**

The first significant organisms from this blood culture (CULTURE-1) will be referred to as:

----- ORGANISM - 1 -----

(9) Enter the identity of ORGANISM-1:

**\*\*UNKNOWN** *{typically the identity is not known and must be inferred}*

(10) Is ORGANISM-1 a rod or coccus (etc.):

**\*\*ROD**

(11) The gram stain of ORGANISM-1:

**\*\*GRAMNEG**

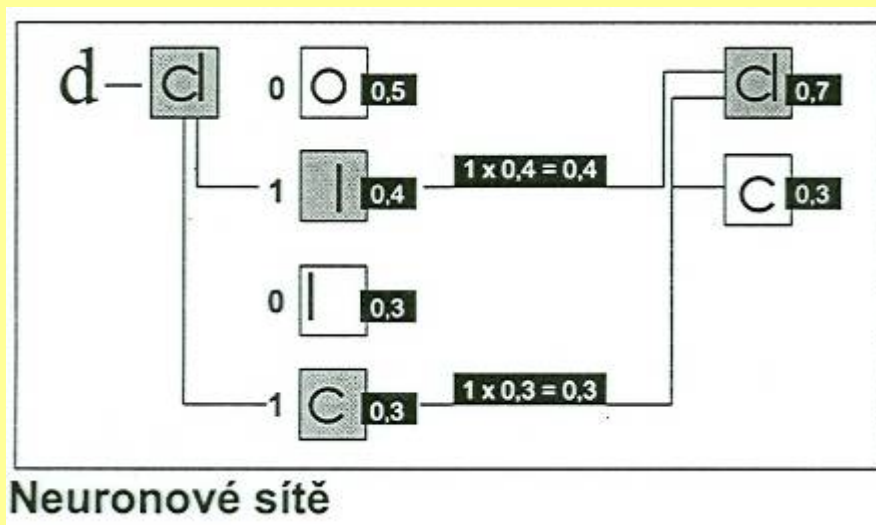
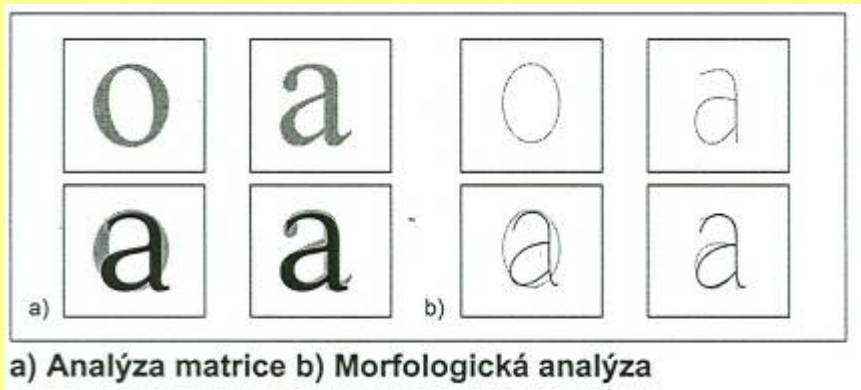
(12) Have there been POSITIVE cultures from sites which are NOT associated with infections about which you are seeking advice?

**\*\*NO**

# Systemy pro analýzy a interpretaci textu

## Automatický zápis textu

OCR – Optical Character Recognition





# Přirozený zdravotnický jazyk

Müllerová, O., Hoffmannová, J., Schneiderová, E.: (1995) *Mluvená čeština v autentických textech*

D pojte si sednout↑ ■ vod koho vy dete↓. ☺ ■■ já vůbec ■ nic nevím↓ ■■

P pane doktor ■ já měla jako sepsi organismu poněvač mi prasklo ■ jako slepý střevo↓ ■■

D nojo↑ no↑ tak↓ ■■ a voni vám teďkon ■ měli říct↑ že ■ že teda tamto já tady nemám ty nějaký detaily ale ...

P já sem to měla zhnisaný tak už štrnáct dní↑ no a nák to mokvá↑ a ■

D ■ nic mě z tohodlenctoho nenapadlo↓ ■ co by to mohlo bejt↓ ■■ nezhoršuje se to po tom prášku↓ ☺

P né↓ akorát tady že se mi to ■ nějak rozlejšá

D helejte se↓ já vám na todlecto↑ napíšu teda tetracyklin

P =hm↑ ■

P pacient, D doktor, ■ pauza, ↓ klesnutí melodie, ↑ stoupnutí melodie, ☺ otázka

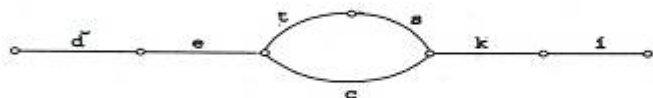


# Systémy pro analýzy a interpretaci řeči

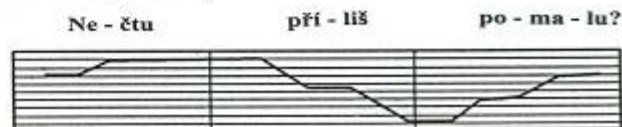
## Mluvená řeč

### 1. Syntéza – pravidla

- **Výslovnost souhlásek**
  - podle následující samohlásky
  - vložená slabá samohláska

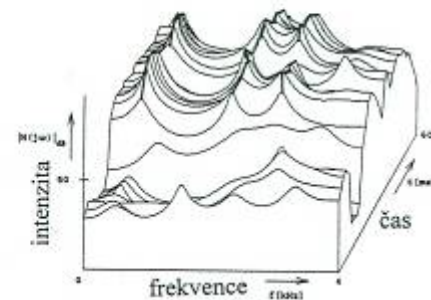


- **Další faktory:** přízvuk, melodie, pauzy



### 2. Analýza – řeči

- **Hranice slov** (pro ženu, proženu) výslovnost stejná, dělení dle kontextu
- **identifikace samohlásek** podle kombinace frekvenčních charakteristik (formantů)



- **identifikace souhlásek**

### Využití analýzy řeči v lékařství

- **hlasové příkazy** – počítač realizuje předem definovanou akci  
př.: volba adresáře
- **diktování**
  - omezený seznam termínů  
př.: sestavení odborného nálezu
  - speaker dependentní systémy – vyžadují fázi učení
- **rozpoznávání volné řeči** – syntaktická a sémantická analýza