

# Formální konceptuální analýza

## moderní metoda analýzy dat

Tomáš Kühn

14. října 2011

# Osnova

- 1 Informatika
- 2 Formální konceptuální analýza
- 3 Matematické pozadí FKA
- 4 Využití FKA

# Co je to informatika?

# Co je to informatika?

“Computer science is no more about computers than astronomy is about telescopes. ”

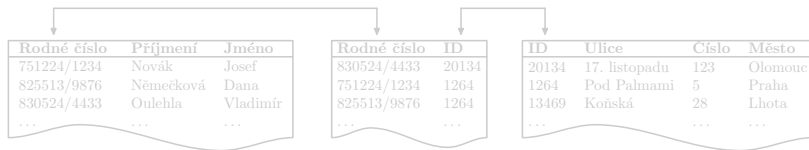
“Informatika se nezabývá počítači o nic více než astronomie dalekohledy. ”

Edsger Dijkstra

# Informatika

Informatika je věda, která se zabývá informacemi (daty) a jejich zpracováním.

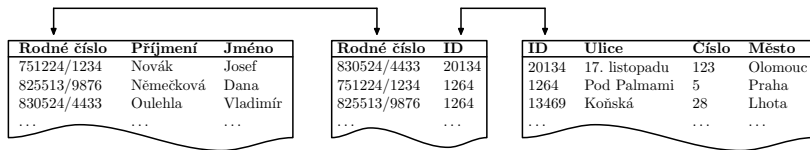
Data nejběžněji zaznamenáváme ve formě vzájemně propojených tabulek (databází).



# Informatika

Informatika je věda, která se zabývá informacemi (daty) a jejich zpracováním.

Data nejběžněji zaznamenáváme ve formě vzájemně propojených tabulek (databází).



# Binární data

- Jsou nejjednodušším typem informace.
- Vyjadřují, zda zkoumaný objekt má či nemá danou vlastnost.
- Jiné typy dat lze relativně snadno převést na binární data.

	2 nohy	4 nohy	vejce	mléko	maso	vlna
pes		×				
kočka		×				
ovce		×		×	×	×
koza		×		×	×	
slepice	×		×		×	
krocan	×				×	

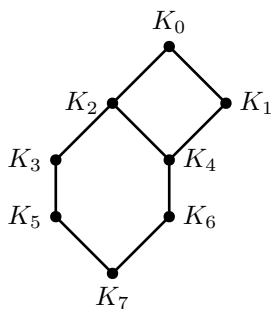
# Binární data

- Jsou nejjednodušším typem informace.
- Vyjadřují, zda zkoumaný objekt má či nemá danou vlastnost.
- Jiné typy dat lze relativně snadno převést na binární data.

	<b>2 nohy</b>	<b>4 nohy</b>	<b>vejce</b>	<b>mléko</b>	<b>maso</b>	<b>vlna</b>
<b>pes</b>		×				
<b>kočka</b>		×				
<b>ovce</b>		×		×	×	×
<b>koza</b>		×		×	×	
<b>slepice</b>	×		×		×	
<b>krocان</b>	×				×	

# Formální konceptuální analýza (FKA)

- metoda analýzy dat,
- jejíž základy položil Rudolf Wille v roce 1982.
- Umožňuje jiný pohled na data.
- Z tabulky s binárními daty je vytvořena množina tzv. konceptů (maximální shluky = obdélníky křížků v tabulce).
- Koncepty lze přirozeně uspořádat.



# Hledání konceptů

Nalezené:

	<b>2 nohy</b>	<b>4 nohy</b>	<b>vejce</b>	<b>mléko</b>	<b>maso</b>	<b>vlna</b>
<b>pes</b>		×				
<b>kočka</b>		×				
<b>ovce</b>		×		×	×	×
<b>koza</b>		×		×	×	
<b>slepice</b>	×		×		×	
<b>krocán</b>	×				×	

## Hledání konceptů

Nalezené:  $\langle \{\text{pes, kočka, ovce, koza}\}, \{4 \text{ nohy}\} \rangle$ ,

	2 nohy	4 nohy	vejce	mléko	maso	vlna
<b>pes</b>		×				
<b>kočka</b>		×				
<b>ovce</b>		×		×	×	×
<b>koza</b>		×		×	×	
<b>slepice</b>	×		×		×	
<b>krocán</b>	×				×	

## Hledání konceptů

Nalezené:  $\langle \{ \text{pes, kočka, ovce, koza} \}, \{ 4 \text{ nohy} \} \rangle$ ,  
 $\langle \{ \text{ovce, koza, slepice, krocán} \}, \{ \text{maso} \} \rangle$ ,

	2 nohy	4 nohy	vejce	mléko	maso	vlna
<b>pes</b>		×				
<b>kočka</b>		×				
<b>ovce</b>		×		×	×	×
<b>koza</b>		×		×	×	
<b>slepice</b>	×		×		×	
<b>krocán</b>	×				×	

## Hledání konceptů

Nalezené:  $\langle \{\text{pes, kočka, ovce, koza}\}, \{4 \text{ nohy}\} \rangle$ ,  
 $\langle \{\text{ovce, koza, slepice, krocan}\}, \{\text{maso}\} \rangle$ ,  
 $\langle \{\text{slepice, krocan}\}, \{2 \text{ nohy, maso}\} \rangle$ ,

	2 nohy	4 nohy	vejce	mléko	maso	vlna
<b>pes</b>		×				
<b>kočka</b>		×				
<b>ovce</b>		×		×	×	×
<b>koza</b>		×		×	×	
<b>slepice</b>	×		×		×	
<b>krocan</b>	×				×	

## Hledání konceptů

Nalezené:  $\langle \{\text{pes, kočka, ovce, koza}\}, \{4 \text{ nohy}\} \rangle$ ,  
 $\langle \{\text{ovce, koza, slepice, krocan}\}, \{\text{maso}\} \rangle$ ,  
 $\langle \{\text{slepice, krocan}\}, \{2 \text{ nohy, maso}\} \rangle$ ,  
 $\langle \{\text{ovce, koza}\}, \{4 \text{ nohy, mléko, maso}\} \rangle$ ,

	2 nohy	4 nohy	vejce	mléko	maso	vlna
<b>pes</b>		×				
<b>kočka</b>		×				
<b>ovce</b>		×		×	×	×
<b>koza</b>		×		×	×	
<b>slepice</b>	×		×		×	
<b>krocan</b>	×				×	

## Hledání konceptů

Nalezené:  $\langle \{ \text{pes, kočka, ovce, koza} \}, \{ 4 \text{ nohy} \} \rangle$ ,  
 $\langle \{ \text{ovce, koza, slepice, krocan} \}, \{ \text{maso} \} \rangle$ ,  
 $\langle \{ \text{slepice, krocan} \}, \{ 2 \text{ nohy, maso} \} \rangle$ ,  
 $\langle \{ \text{ovce, koza} \}, \{ 4 \text{ nohy, mléko, maso} \} \rangle$ ,  
 $\langle \{ \text{slepice} \}, \{ 2 \text{ nohy, vejce, maso} \} \rangle$ ,

	2 nohy	4 nohy	vejce	mléko	maso	vlna
<b>pes</b>		×				
<b>kočka</b>		×				
<b>ovce</b>		×		×	×	×
<b>koza</b>		×		×	×	
<b>slepice</b>	×		×		×	
<b>krocan</b>	×				×	

## Hledání konceptů

Nalezené:  $\langle \{\text{pes, kočka, ovce, koza}\}, \{4 \text{ nohy}\} \rangle$ ,  
 $\langle \{\text{ovce, koza, slepice, krocan}\}, \{\text{maso}\} \rangle$ ,  
 $\langle \{\text{slepice, krocan}\}, \{2 \text{ nohy, maso}\} \rangle$ ,  
 $\langle \{\text{ovce, koza}\}, \{4 \text{ nohy, mléko, maso}\} \rangle$ ,  
 $\langle \{\text{slepice}\}, \{2 \text{ nohy, vejce, maso}\} \rangle$ ,  
 $\langle \{\text{ovce}\}, \{4 \text{ nohy, mléko, maso, vlna}\} \rangle$ ,

	2 nohy	4 nohy	vejce	mléko	maso	vlna
<b>pes</b>		×				
<b>kočka</b>		×				
<b>ovce</b>		×		×	×	×
<b>koza</b>		×		×	×	
<b>slepice</b>	×		×		×	
<b>krocan</b>	×				×	

## Uspořádání konceptů

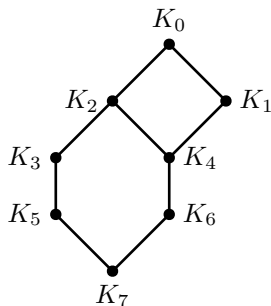
Koncepty:  $K_1 = \langle \{\text{pes, kočka, ovce, koza}\}, \{4 \text{ nohy}\} \rangle$ ,  
 $K_2 = \langle \{\text{ovce, koza, slepice, krocan}\}, \{\text{maso}\} \rangle$ ,  
 $K_3 = \langle \{\text{slepice, krocan}\}, \{2 \text{ nohy, maso}\} \rangle$ ,  
 $K_4 = \langle \{\text{ovce, koza}\}, \{4 \text{ nohy, mléko, maso}\} \rangle$ ,  
 $K_5 = \langle \{\text{slepice}\}, \{2 \text{ nohy, vejce, maso}\} \rangle$ ,  
 $K_6 = \langle \{\text{ovce}\}, \{4 \text{ nohy, mléko, maso, vlna}\} \rangle$ ,  
 $K_0 = \langle \{\text{pes, kočka, ovce, koza, slepice, krocan}\}, \emptyset \rangle$ ,  
 $K_7 = \langle \emptyset, \{2 \text{ nohy, 4 nohy, vejce, mléko, maso, vlna}\} \rangle$ .

Koncepty uspořádáme podle odpovídajících množin objektů:

- např.  $\{\text{ovce, koza}\} \subseteq \{\text{pes, kočka, ovce, koza}\} \Rightarrow K_4 \leq K_1$
- $K_7 \leq K_5 \leq K_3 \leq K_2 \leq K_0, K_7 \leq K_6 \leq K_4 \leq K_1 \leq K_0,$   
 $K_4 \leq K_2.$

## Konceptuální svaz

- Uspořádaná množina konceptů tvoří tzv. konceptuální svaz,
- který může být graficky znázorněn Hasseovým diagramem.
- Jednotlivé koncepty jsou v tomto diagramu znázorněny uzly.
- Koncepty  $K_i$  a  $K_j$  jsou v diagramu spojeny, pokud  $K_i \leq K_j$ , při čemž  $K_j$  je umístěn výše než  $K_i$ .



# Formální kontext

## Definice

*Formální kontext* je trojice  $\langle X, Y, I \rangle$ , kde

- $X$  je množina *objektů*,
- $Y$  je množina *atributů*
- a  $I$  je binární relace *incidence* mezi množinami  $X$  a  $Y$ .

# Formální koncept

## Definice

Pro dané množiny  $A \subseteq X$  a  $B \subseteq Y$  definujeme

- $A^\uparrow = \{y \in Y \mid \forall x \in A \text{ platí } \langle x, y \rangle \in I\}$ ,
- $B^\downarrow = \{x \in X \mid \forall y \in B \text{ platí } \langle x, y \rangle \in I\}$ .

Pokud  $A^\uparrow = B$  a  $B^\downarrow = A$ , pak dvojici  $\langle A, B \rangle$  nazýváme *formální koncept*. Množinu  $A$  nazýváme *extent* a množinu  $B$  *intent* tohoto konceptu.

Pro dva formální koncepty  $\langle A_1, B_1 \rangle$  a  $\langle A_2, B_2 \rangle$  klademe  $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$ , právě když  $A_1 \subseteq A_2$ .

# Využití FKA

- obecně – jiný pohled na data,
- vyhledávání informací,
- na katedře informatiky UP:
  - vyhodnocování dotazníků,
  - taxonomie zkamenělin.

# Literatura

- C. Carpineto, G. Romano: *Concept Data Analysis: Theory and Applications*. John Wiley & Sons, 2004.
- B. Ganter, R. Wille: *Formal Concept Analysis – Mathematical Foundations*. Springer, 1999.
- R. Wille: *Restructuring lattice theory: an approach based on hierarchies of concepts*. *Ordered Sets*, strany 445–470. Boston, 1982.
- Tuto prezentaci je možné stáhnout na adrese <http://www.inf.upol.cz/prednasky-pro-stredni-skoly/>.