



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Reintegrace českého vědce a vytvoření špičkového týmu v informačních vědách

Formální konceptuální analýza

Metoda zpracování dat s filozofickým pozadím

Michal Krupka

Studentský vědecký seminář

6. října 2011

Příklad: snídane

Ingredience

Chléb

Máslo

Vejce

Cibule

Med

Olej

Sladký rohlík

Sůl

Jídla

Chléb s máslem

Vejce na měkko

Volská oka

Míchaná vejce

Míchaná vejce na cibulce

Chléb s máslem a cibulí

Chléb s máslem a medem

Sladký rohlík

Sladký rohlík s máslem

Příklad: snídane

	ch	má	vej	cb	md	ol	sr	s
Chm	x	x						x
Vm	x	x	x					x
Vol	x		x			x		x
Mv	x		x			x		x
Mvc	x		x	x		x		x
Cmc	x	x		x				x
Cmm	x	x			x			
Sr							x	
Srm		x					x	

Otázky

- Jak zjistit jídla, která obsahují dané ingredience?
- Jak zjistit ingredience, které jsou v daných jídlech?
- „Mám chuť na vajíčka“, „mám chuť na chléb s máslem“
- Nemám olej. Jaké jídlo si mohu připravit?

Jídla

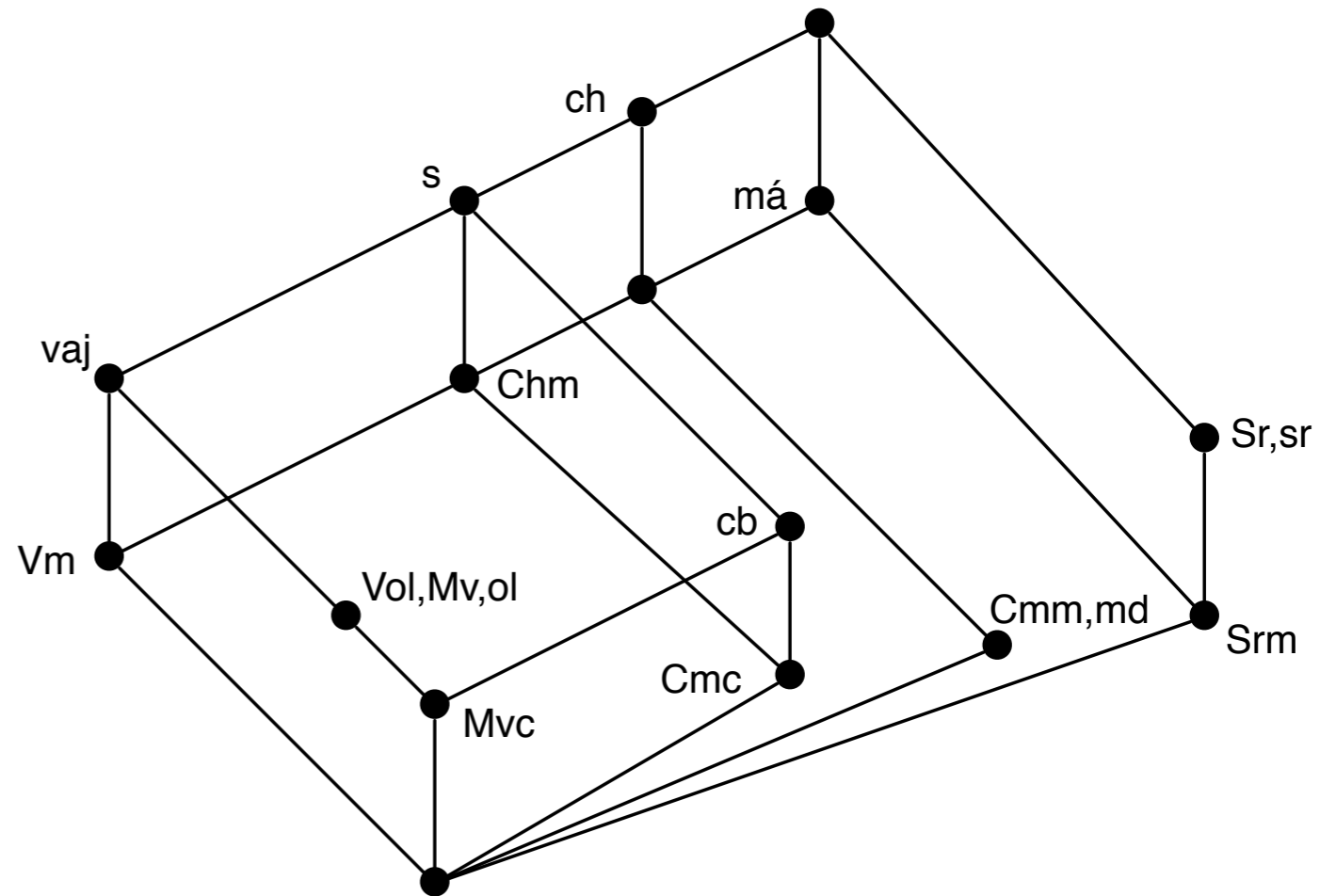
Chm: chléb s máslem, Vm: vejce na měkko, Vol: volská oka, Mv: míchaná vejce, Mvc: míchaná vejce na cibulce, Cmc: chléb s máslem a cibulí, Cmm: chléb s máslem a medem, Sr: sladký rohlík, Srm: sladký rohlík s máslem.

Ingredience

ch: chléb, má: máslo, vej: vejce, cb: cibule, md: med, ol: olej, sr: sladký rohlík, s: sůl

Příklad: snídane

	ch	má	vej	cb	md	ol	sr	s
Chm	x	x						x
Vm	x	x	x					x
Vol	x		x			x		x
Mv	x		x			x		x
Mvc	x		x	x		x		x
Cmc	x	x		x				x
Cmm	x	x			x			
Sr							x	
Srm		x					x	



Jídla

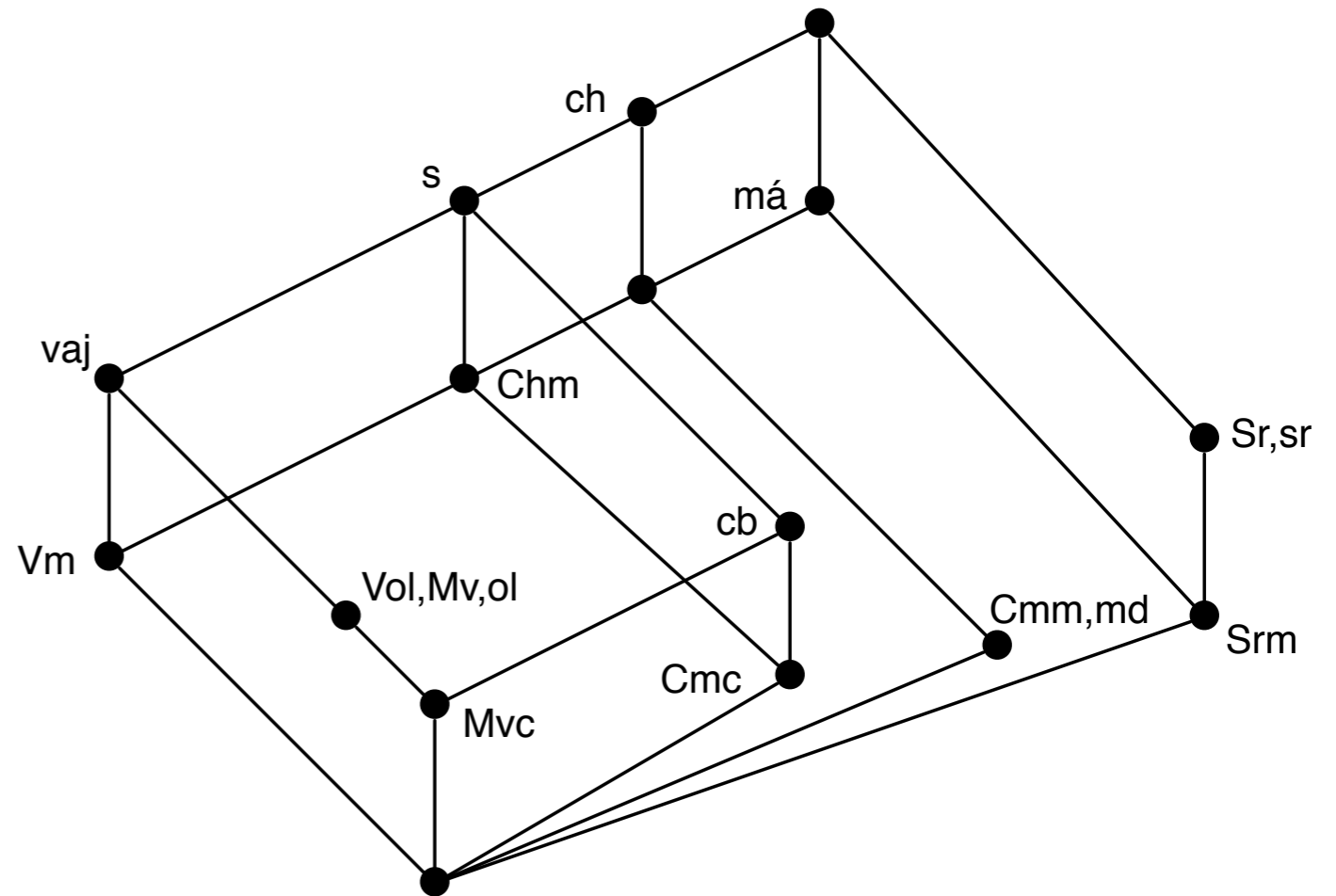
Chm: chléb s máslem, Vm: vejce na měkko, Vol: volská oka, Mv: míchaná vejce, Mvc: míchaná vejce na cibulce, Cmc: chléb s máslem a cibulí, Cmm: chléb s máslem a medem, Sr: sladký rohlík, Srm: sladký rohlík s máslem

Ingredience

ch: chléb, má: máslo, vej: vejce, cb: cibule, md: med, ol: olej, sr: sladký rohlík, s: sůl

Příklad: snídane

	ch	má	vej	cb	md	ol	sr	s
Chm	x	x						x
Vm	x	x	x					x
Vol	x		x			x		x
Mv	x		x			x		x
Mvc	x		x	x		x		x
Cmc	x	x		x				x
Cmm	x	x			x			
Sr							x	
Srm		x					x	



Pozorování

- Ne všechny uzly v obrázku jsou označeny.
- Jak interpretovat uzly v obrázku?
- Jak zkonstruovat obrázek z tabulky?
- Lze zkonstruovat tabulku z obrázku?

Uspořádané množiny

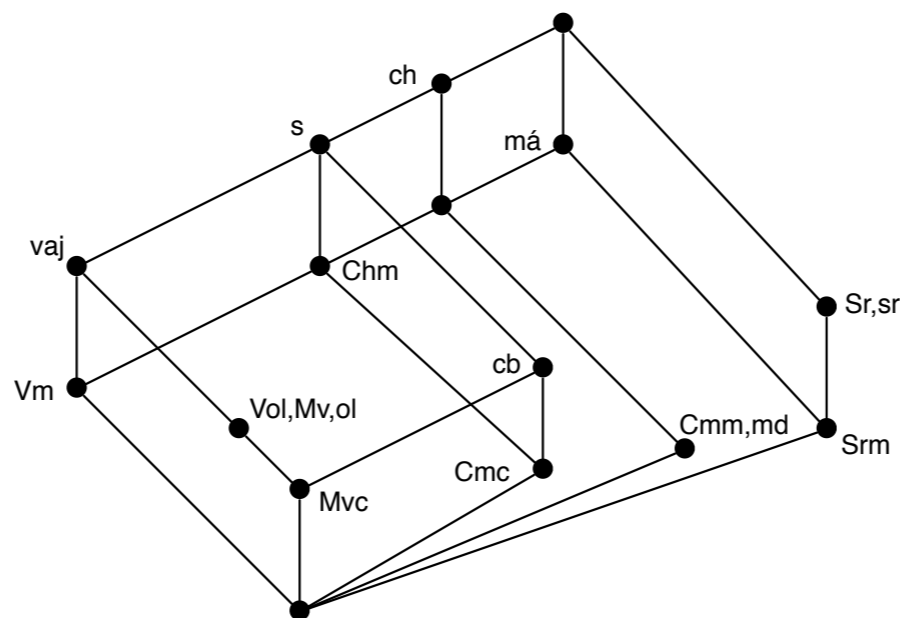
Uspořádání

Binární relace \leq na množině X se nazývá *uspořádání*, jestliže je

- *reflexivní*, tj. pro každé $x \in X$ platí $x \leq x$,
- *antisymetrická*, tj. $x \leq y$ a $y \leq x$ jedině když $x = y$,
- *tranzitivní*, tj. z $x \leq y$ a $y \leq z$ plyne $x \leq z$.

Množina X s uspořádáním \leq se nazývá *uspořádaná množina*.

Zobrazení uspořádané množiny Hasseovým diagramem



Úplné svazy

Horní a dolní závora

Nechť X je uspořádaná množina, $Y \subseteq X$. Prvek $x \in X$ se nazývá *horní závora množiny* Y , jestliže pro každé $y \in Y$ platí $y \leq x$. Prvek $x \in X$ se nazývá *dolní závora množiny* Y , jestliže pro každé $y \in Y$ platí $x \leq y$.

Supremum a infimum

Prvek $x \in X$ se nazývá *supremum* množiny Y , jestliže je její horní závorou a pro každou další její horní závoru z platí $x \leq z$ (x je nejmenší horní závora Y). Píšeme $x = \bigvee Y$.

Prvek $x \in X$ se nazývá *infimum* množiny Y , jestliže je její dolní závorou a pro každou další její dolní závoru z platí $x \geq z$ (x je největší dolní závora Y). Píšeme $x = \bigwedge Y$.

Úplný svaz

Uspořádaná množina X se nazývá *úplný svaz*, jestliže každá její podmnožina má supremum a infimum.

Konceptuální svazy

Formální kontext

Formální kontext je trojice $\langle X, Y, I \rangle$, kde X je množina objektů, Y množina atributů a I binární relace mezi X a Y .

Formální koncept Pro množiny $A \subseteq X$ a $B \subseteq Y$ klademe

$$A^\uparrow = \{y \in Y \mid \text{pro každé } x \in A \text{ platí } \langle x, y \rangle \in I\}$$

$$B^\downarrow = \{x \in X \mid \text{pro každé } y \in B \text{ platí } \langle x, y \rangle \in I\}$$

Pokud $A^\uparrow = B$ a $B^\downarrow = A$, nazýváme dvojici $\langle A, B \rangle$ *formální koncept*, množina A se nazývá *extent* a množina B *intent* tohoto formálního konceptu.

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

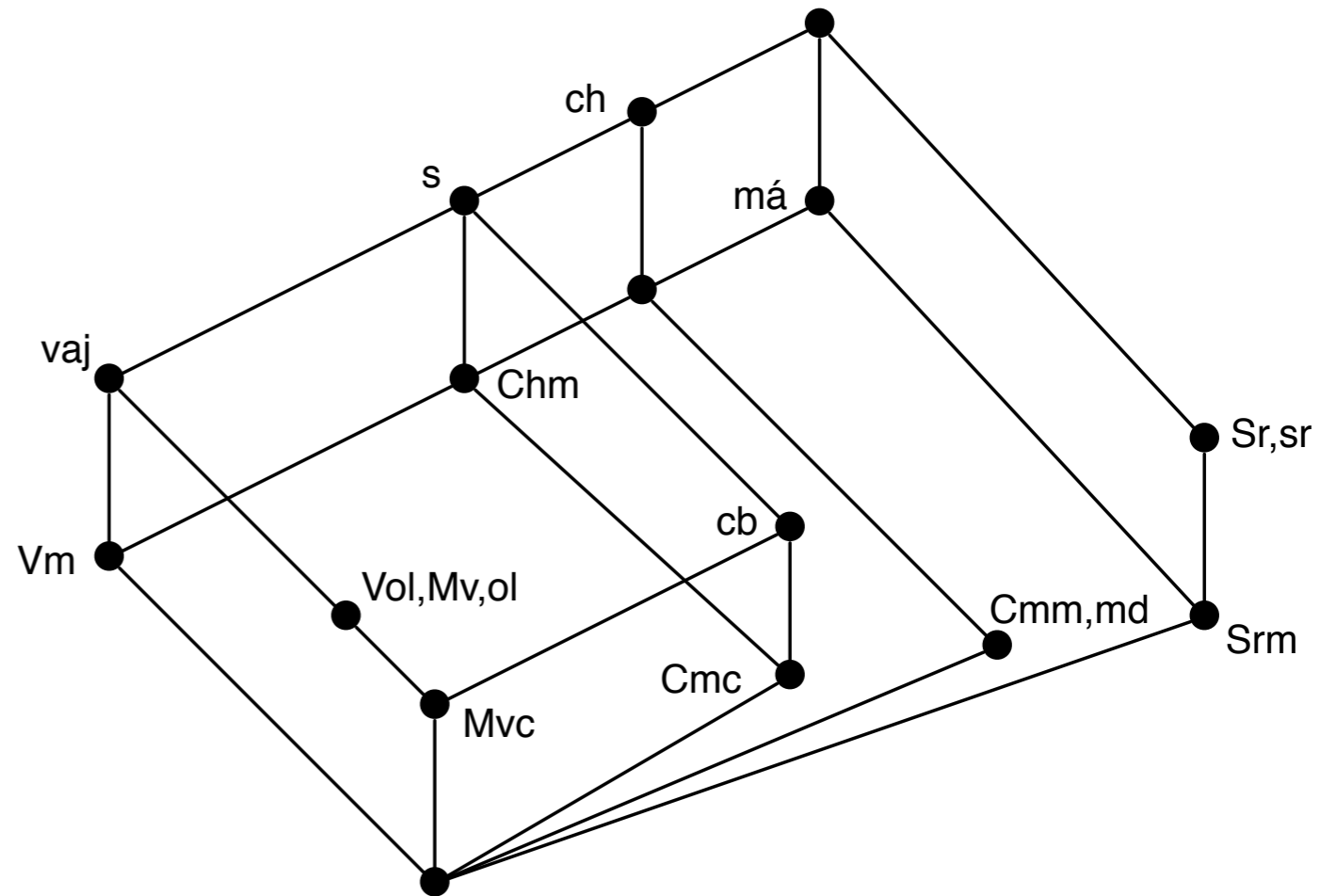
Základní výsledek

Množina $\mathcal{B}(X, Y, I)$ všech formálních konceptů formálního kontextu $\langle X, Y, I \rangle$ spolu s uspořádáním \leq je úplný svaz.

Tento úplný svaz se nazývá *konceptuální svaz*

Zpět k příkladu

	ch	má	vej	cb	md	ol	sr	s
Chm	x	x						x
Vm	x	x	x					x
Vol	x		x			x		x
Mv	x		x			x		x
Mvc	x		x	x		x		x
Cmc	x	x		x				x
Cmm	x	x			x			
Sr							x	
Srm		x					x	



Jídla

Chm: chléb s máslem, Vm: vejce na měkko, Vol: volská oka, Mv: míchaná vejce, Mvc: míchaná vejce na cibulce, Cmc: chléb s máslem a cibulí, Cmm: chléb s máslem a medem, Sr: sladký rohlík, Srm: sladký rohlík s máslem

Ingredience

ch: chléb, má: máslo, vej: vejce, cb: cibule, md: med, ol: olej, sr: sladký rohlík, s: sůl

Příklady aplikací

- obecně: jiný pohled na data
 - vyhledávání informací (information retrieval)
 - analýza zdrojového kódu programu
 - vyhodnocování dotazníků
 - taxonomie zkamenělin
-
- další aplikace viz např. kniha [1]

Kde je to filozofické pozadí?

- v širším slova smyslu: zpřístupnění abstraktní matematické teorie uživateli [3]
- v užším slova smyslu: redukce konceptu na extent a intent [3, port-royalská logika]

Poznámky

- toto je jen začátek: k řešení problému obvykle nestačí pouze zkonstruovat konceptuální svaz
- hlavní potíže: velikost konceptuálního svazu

Úloha

- napište co nejrychlejší program na vygenerování konceptuálního svazu (podrobnosti na webu)

Literatura

- [1] C. Carpineto and G. Romano. *Concept Data Analysis: Theory and Applications*. John Wiley & Sons, 2004.
- [2] B. Ganter and R. Wille. *Formal Concept Analysis – Mathematical Foundations*. Springer, 1999.
- [3] R. Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. In I. Rival, editor, *Ordered Sets*, pages 445–470. Boston, 1982. seminal publication on Formal Concept Analysis.