Konceptuální svazy

Daniela Ďuráková

Katedra informatiky, FEI, VŠB-TUO
daniela.durakova@vsb.cz

Proč použití hierarchických struktur?

- výsledek dotazu - možné zahlcení uživatele množstvím vyhovujících objektů

- uspořádání objektů
- ohodnocení objektů mírou jejich příslušnosti ke sledovaným vlastnostem

Dotaz

- sada parametrů
- vliv rozsahu každého z parametrů na náležitost každého objektu z hlediska kvality objektu
- výsledek - vlastnost objekt má nebo nemá
- výsledek - objekt vlastnost má v určité míře vyjádřené numerickou hodnotou

Výsledek dotazu s boolean hodnotami

rostlina	bílá	modrá	žlutá	červená	zelená
kopretina	x		x		x
zvonek		x			x
vlčí mák				x	x
sněženka	x				x
něco s červenými listy		x		x	

Teoretické základy

- použité pojmy

- *kontext, koncept*
- *Galoisova konexe*
- *teorie svazů*
- *fuzzy teorie, α - řezy*
- *uspořádané množiny*

Kontext

Definice

Formální kontext $\boldsymbol{K = (G, M, I)}$ se skládá z množin $\boldsymbol{G}$ a $\boldsymbol{M}$ a relace $\boldsymbol{I} \subseteq \boldsymbol{G} \times \boldsymbol{M}$.

Prvky množiny $\boldsymbol{G}$ se nazývají *objekty* a prvky množiny $\boldsymbol{M}$ se nazývají *atributy*.
Řekneme, že objekt $\boldsymbol{g}$ má atribut $\boldsymbol{m}$, jestliže $\boldsymbol{(g, m)} \in \boldsymbol{I}$. Znamená to, že $\boldsymbol{g}$ je v relaci s $\boldsymbol{m}$, píšeme $\boldsymbol{gIm}$.

Příklad kontextu

G = {kopretina, zvonek, vlčí mák, ...}, M = {barvy rostlin}

rostlina	bílá	modrá	žlutá	červená	zelená
kopretina	x		x		x
zvonek		x			x
vlčí mák				x	x
sněženka	x				x
něco s červenými listy		x		x	

Galoisova konexe

Nechť $\boldsymbol{A} \subseteq \mathbf{G}$, $\boldsymbol{B} \subseteq \boldsymbol{M}$. ***Galoisovu konexi*** tvoří (φ, σ), pro které platí

$\varphi(\boldsymbol{A}) = \{m \in \boldsymbol{M};\ g\boldsymbol{I}m \text{ pro} \forall\, g \in \boldsymbol{A}\}$

to znamená - množina všech atributů společných pro všechny objekty z ***A***

$\sigma(\boldsymbol{B}) = \{g \in \boldsymbol{G};\ g\boldsymbol{I}m \text{ pro } \forall\ m \in \boldsymbol{B}\}$

to znamená - množina všech objektů, které mají všechny atributy z ***B***

Formální koncept

Definice

Formální koncept kontextu $\boldsymbol{K = (G, M, I)}$ je dvojice $\boldsymbol{(A, B)}$ kde $\boldsymbol{A \subseteq G}$, $\boldsymbol{B \subseteq M}$

a platí $\varphi\boldsymbol{(A) = B}$ a $\sigma\boldsymbol{(B) = A}$.

ß(G, M, I) označíme množinu všech konceptů kontextu ***K*.**

Co je koncept?

rostlina	bílá	žlutá	modrá	červená	zelená
kopretina	x	x			x
zvonek			x		x
vlčí mák				x	x
sněženka	x				x
blázen			x	x	

Co není koncept?

rostlina	bílá	žlutá	modrá	červená	zelená
kopretina	x	x			x
zvonek			x		x
vlčí mák				x	x
sněženka	x				x
blázen			x	x	

Uspořádání konceptů

Mějme (A_1, B_1), (A_2, B_2) kde $A_1, A_2 \subseteq G$, $B_1, B_2 \subseteq M$ a platí $A_1 \subseteq A_2$, $B_2 \subseteq B_1$, pak *koncept* (A_1, B_1) nazveme ***podkonceptem*** konceptu (A_2, B_2), *koncept* (A_2, B_2) nazveme ***nadkonceptem*** konceptu (A_1, B_1).

Píšeme $(A_1, B_1) \leq (A_2, B_2)$.

Konceptuální svaz

- *Minimem* uspořádání je koncept označovaný ***bot***, který je tvořen dvojicí množin – množinou objektů se všemi atributy a množinou všech atributů.
- *Maximem* uspořádání je koncept označovaný ***top***, který je tvořen dvojicí množin – množinou atributů pro všechny objekty a množinou všech objektů.
- Uspořádané koncepty tvoří ***svaz***, který je označován jako ***konceptuální svaz.***

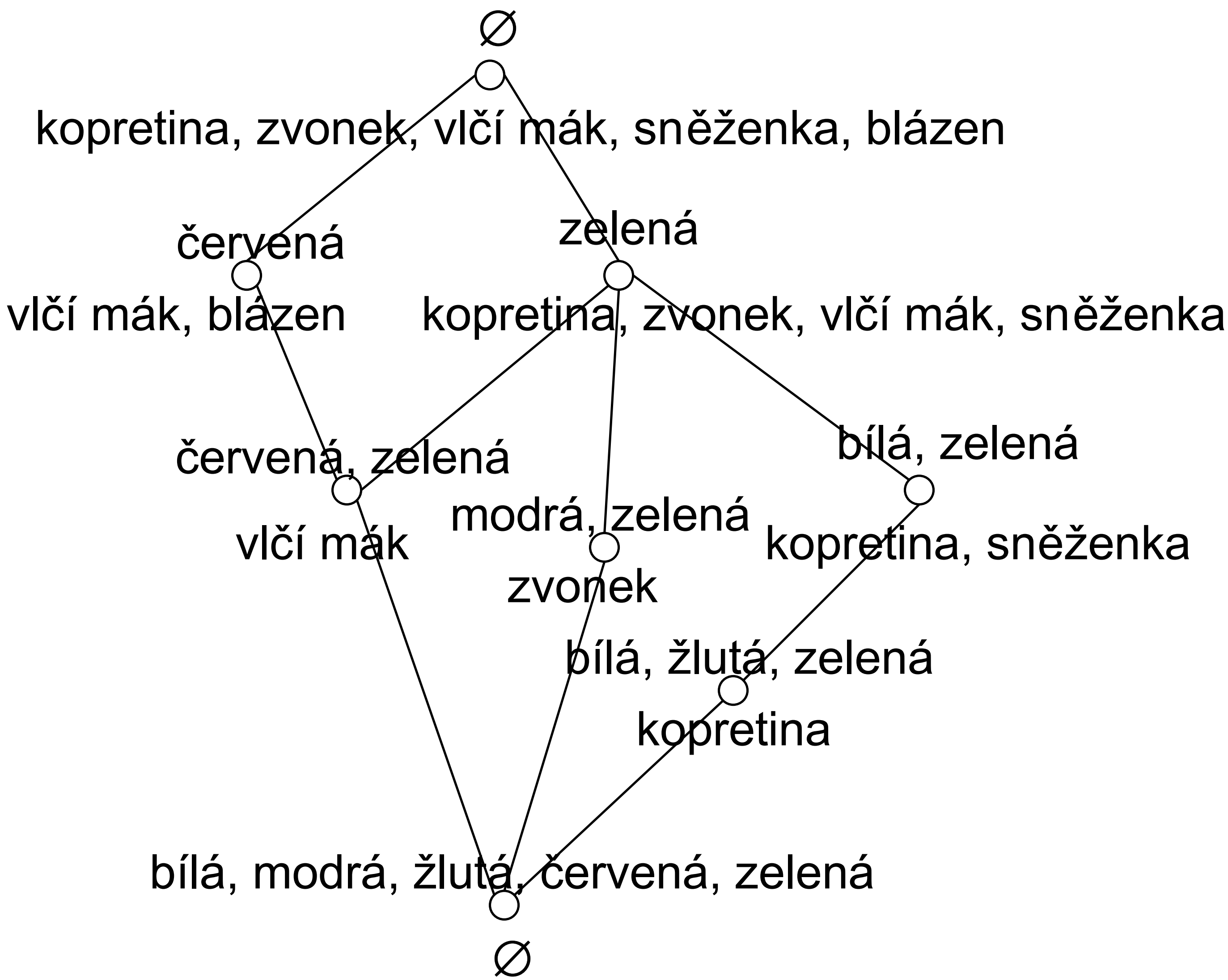
Kontextová tabulka

rostlina	bílá	modrá	žlutá	červená	zelená
kopretina	x		x		x
zvonek		x			x
vlčí mák				x	x
sněženka	x				x
něco s červenými listy		x		x	

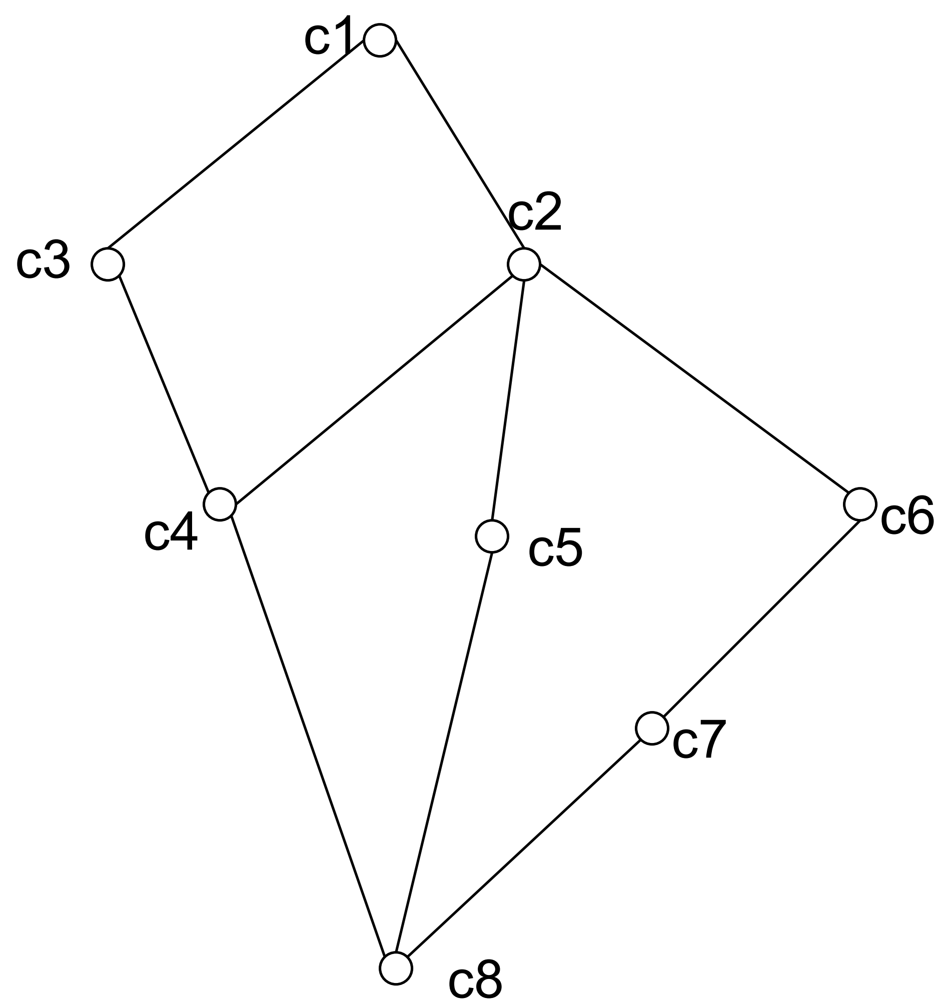
Koncepty (květiny)

koncepty	kopre tina	zvo- nek	mák	sně- ženka	blá- zen					
c1	x	x	x	x	x					
c2	x	x	x	x						x
c3			x		x				x	
c4			x						x	x
c5		x					x			x
c6	x			x		x				x
c7	x					x		x		x
c8						x	x	x	x	x

Konceptuální svaz (květiny)



Konceptuální svaz (květiny)



Dotaz s výsledkem v num. hodnotách

- s parametry

- vzdálenost z Prahy < 500 km (**d**)
- cena skipasu < 5200 Kč (**s**)
- nadmořská výška 1500 metrů n.m. (**e**)

Výsledek dotazu s num. hodnotami

středisko	zkratka	d	s	e
Mayrhofen	Ma	483	5276	3250
Sölden	So	576	4866	3260
Kitzbühel	Ki	465	4741	2000
Flattach	Fl	490	4411	3125
Söll	Sl	460	3664	1835
Zell am See	Ze	482	4632	3029
Radstadt	Ra	450	4625	2130
Gosau	Go	390	3774	1600
Rohrmoos	Ro	426	4565	2700

Škálování kontextu

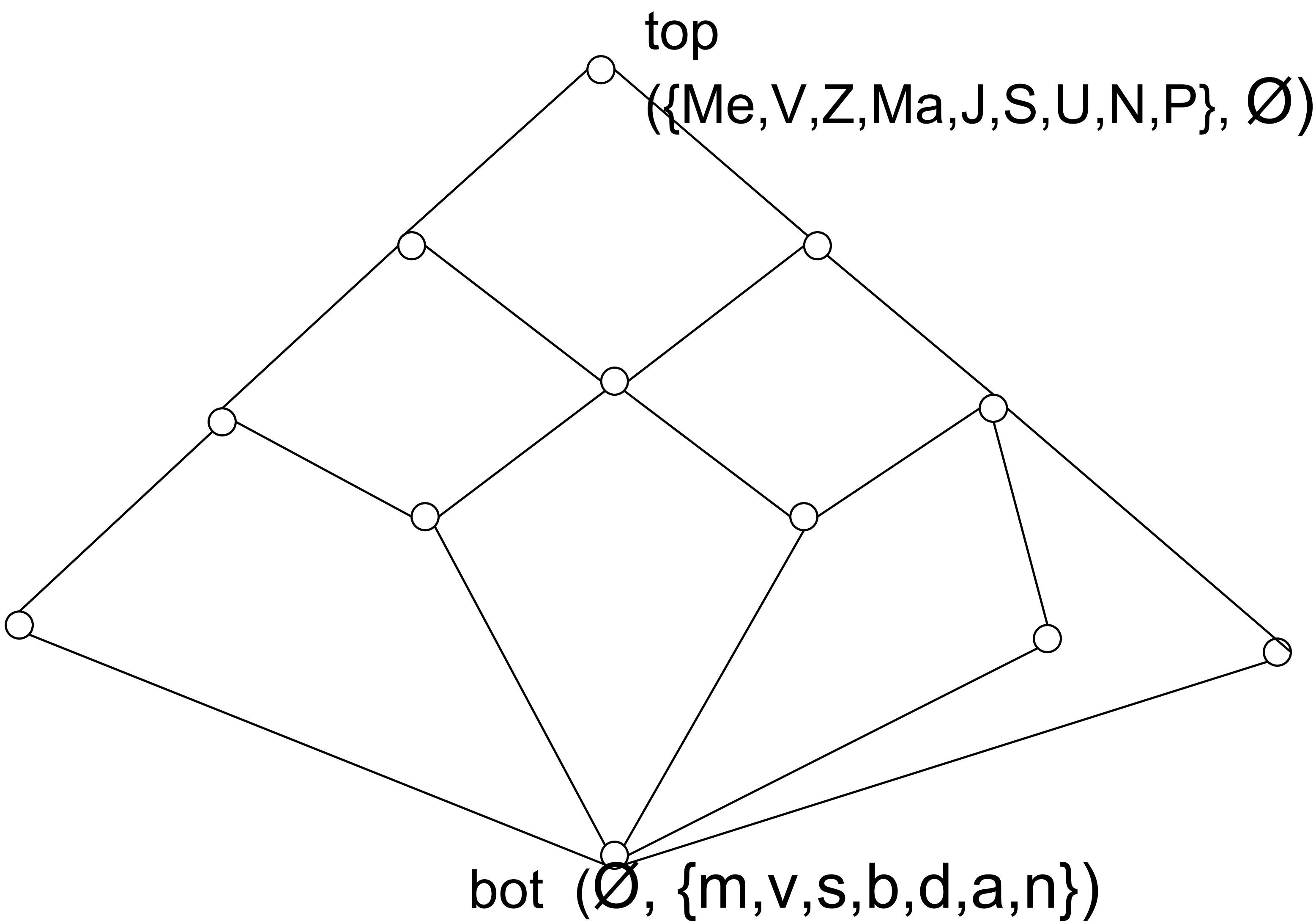
- nominální škálování
- ordinální škálování

- B. Ganter, R. Wille: *Formal Concept Analysis,* Springer Verlag 1999.

Příklad kontextu (nominální škálování)

	velikost			od Slunce		měsíc	
	malá	střední	velká	blízko	daleko	ano	ne
Merkur	x			x			x
Venuše	x			x			x
Země	x			x		x	
Mars	x			x		x	
Jupiter			x		x	x	
Saturn			x		x	x	
Uran		x			x	x	
Neptun		x			x	x	
Pluto	x				x	x	

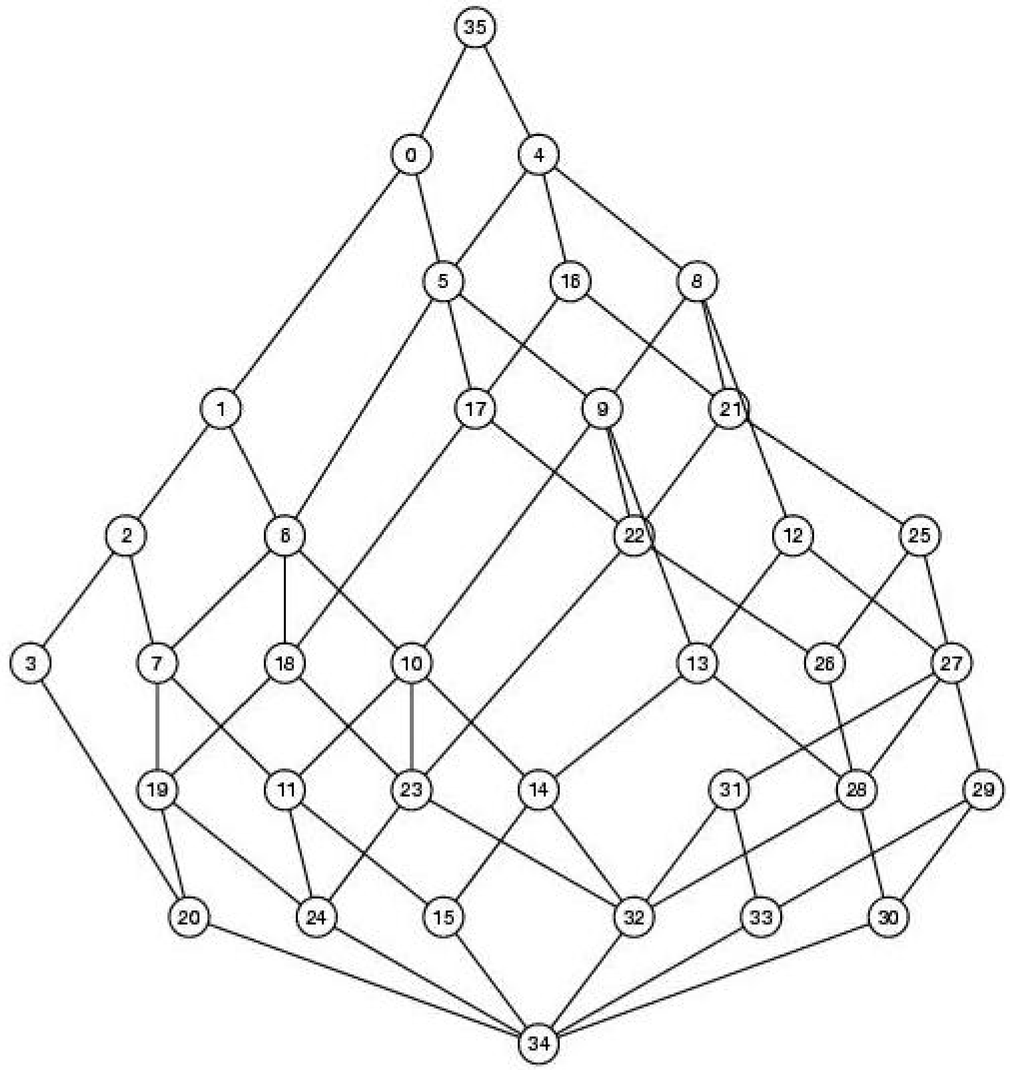
Konceptuální svaz (planety)



Tvorba kontextu pomocí škálování

	vzdálenost ≤				cena tis. Kč ≤				výška tis. n.m.>			
	550	490	470	440	4.8	4.5	4.4	4.0	2.0	2.5	3.0	3.2
Ma	x	x							x	x	x	x
So									x	x	x	x
Ki	x	x	x		x							
Fl	x				x	x			x	x	x	
Sl	x	x	x		x	x	x		x			
Ze	x	x			x				x	x	x	
Ra	x	x	x		x				x			
Go	x	x	x	x	x	x	x	x				
Ro	x	x	x	x	x	x			x	x		

Konceptuální svaz (škálovaný kontext)



Využití fuzzy teorie

- fuzzy příslušnost atributů - proložením lineární funkce (hraniční body dány parametry dotazu)
- fuzzy kontextová tabulka
- Radim Bělohlávek
- *Similarity relations in concept lattices* (Journal of Logic and Computation) Vol.10 No. 6(2000), 823-845. [Oxford University Press]
- *Fuzzy Galois connections.* (Math. Logic Quarterly 45),4 (1999), 497-504.[Wiley-VCH, ISSN 0942-5616]

Problémy

- velký nárůst počtu konceptů
- závislost na zvolené škále

- snaha odstranit obojí
- snížení počtu konceptů - normalizace kontextu
- snížení míry subjektivity při volbě škály

Fuzzy kontext

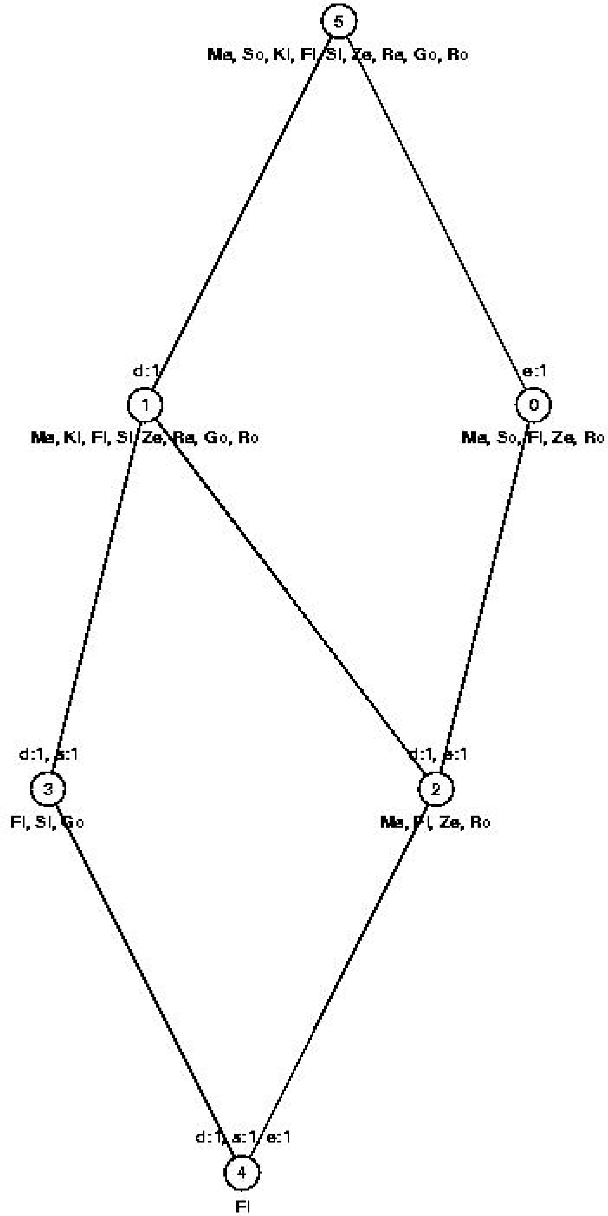
středisko	v	s	e
Mayrhofen	0.58	0.00	0.85
Sölden	0.12	0.09	0.86
Kitzbühel	0.67	0.17	0.12
Flattach	0.55	0.39	0.78
Söll	0.70	0.89	0.02
Zell am See	0.59	0.24	0.72
Radstadt	0.75	0.25	0.19
Gosau	1.00	0.82	0.00
Rohrmoos	0.87	0.29	0.53

α-řezy

- příslušnost atributů v požadované míře
- je-li K fuzzy kontext, pak pro $\alpha \in [0, 1]$ je
  ${}^{\alpha}K = \{x \in X, K(x) \geq \alpha\}$
- pro každý α-kontext nalezneme koncepty
- sjednocení těchto konceptů nazveme
  α- koncept, $\boldsymbol{c}_\alpha = \{G(\boldsymbol{c}_\alpha), M(\boldsymbol{c}_\alpha)\}$
- každému prvku množiny $M(\boldsymbol{c}_\alpha)$ každého α-konceptu přiřadíme váhu, která je dána počtem jeho opakování v ***α-uspořádané struktuře***

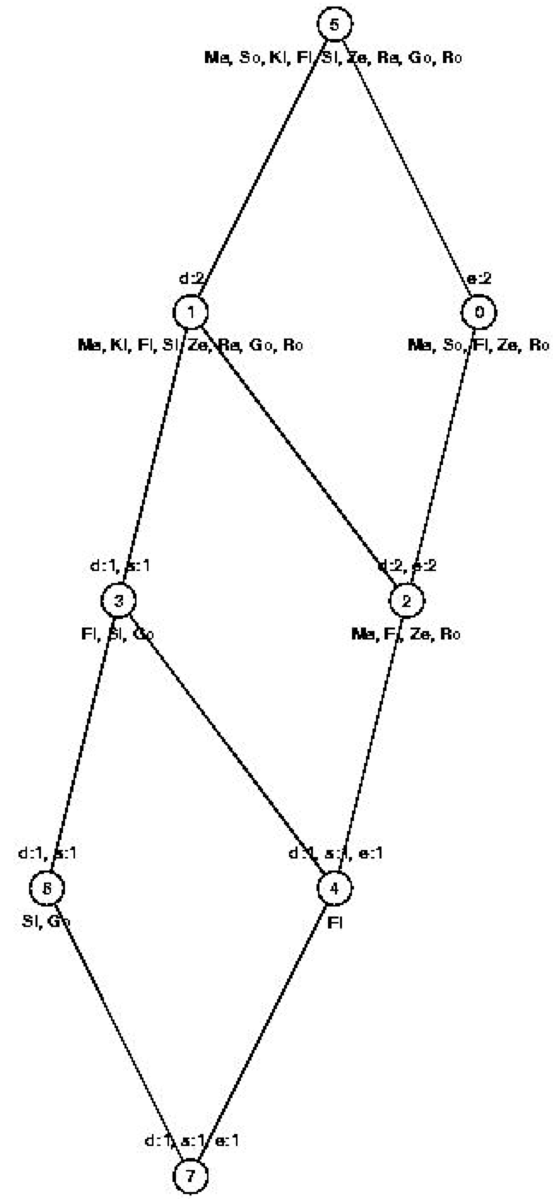
α-řez pro α = 0.4

	d	s	e
Ma	x		x
So			x
Ki	x		
Fl	x		x
Sl	x	x	
Ze	x		x
Ra	x		
Go	x	x	
Ro	x		x



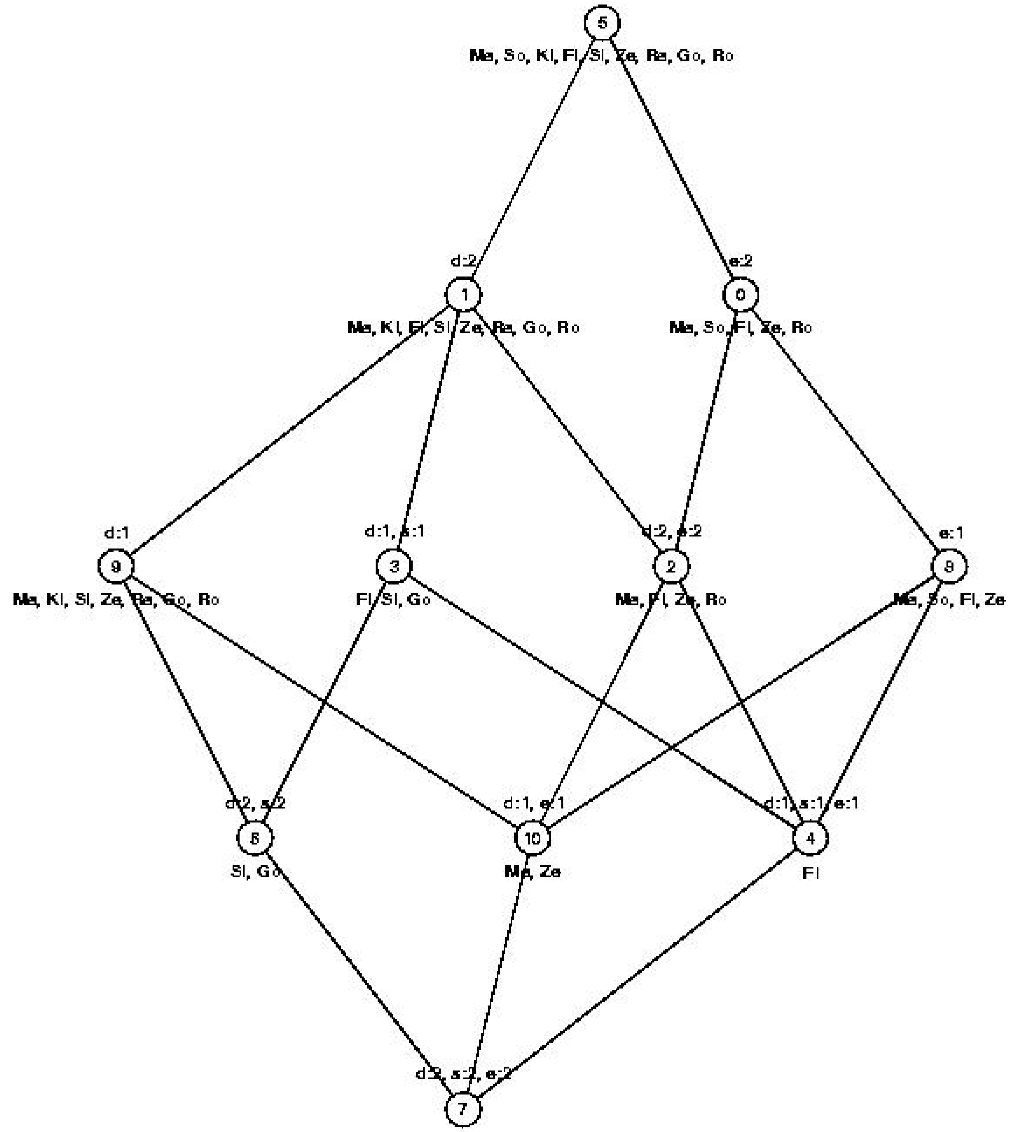
α-řez pro α = 0.5

	d	s	e
Ma	x		x
So			x
Ki	x		
Fl	x		x
Sl	x	x	
Ze	x		x
Ra	x		
Go	x	x	
Ro	x		x



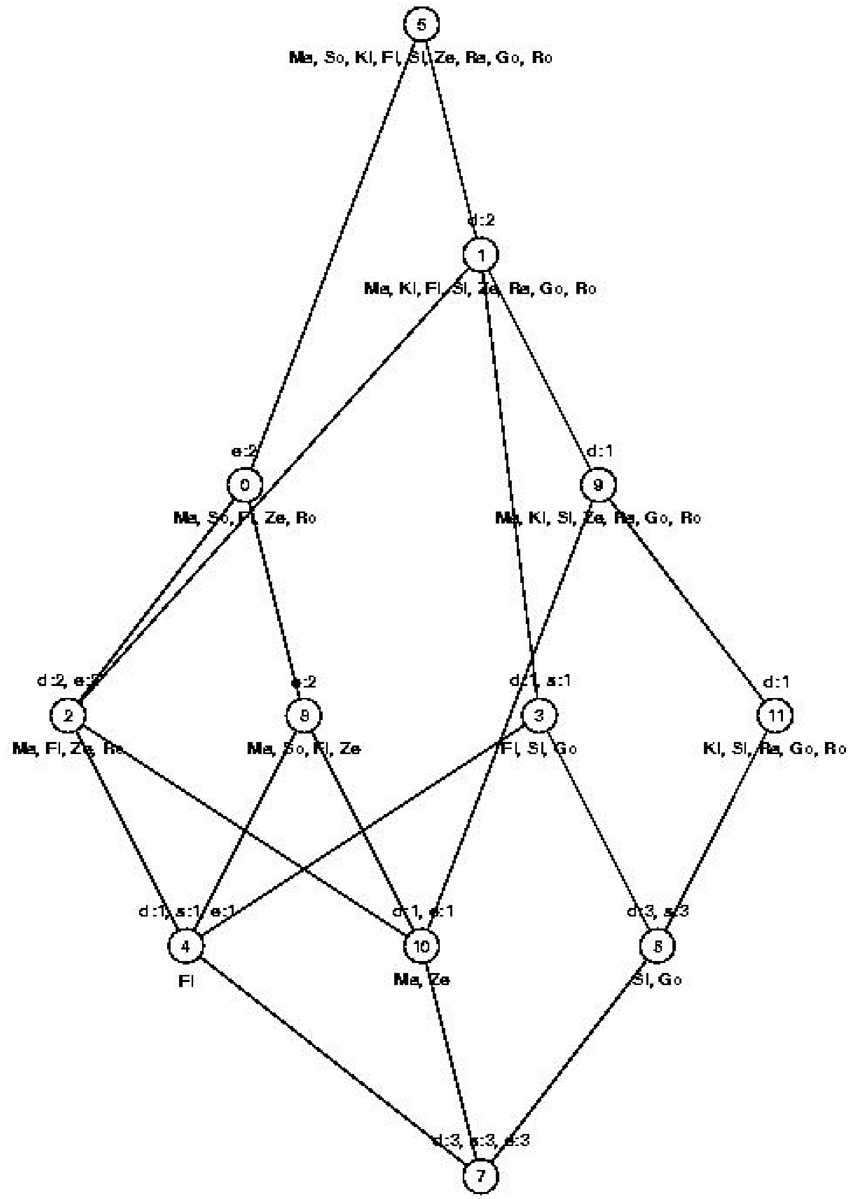
α-řez pro α = 0.6

	d	s	e
Ma	x		x
So			x
Ki	x		
Fl			x
Sl	x	x	
Ze	x		x
Ra	x		
Go	x	x	
Ro	x		



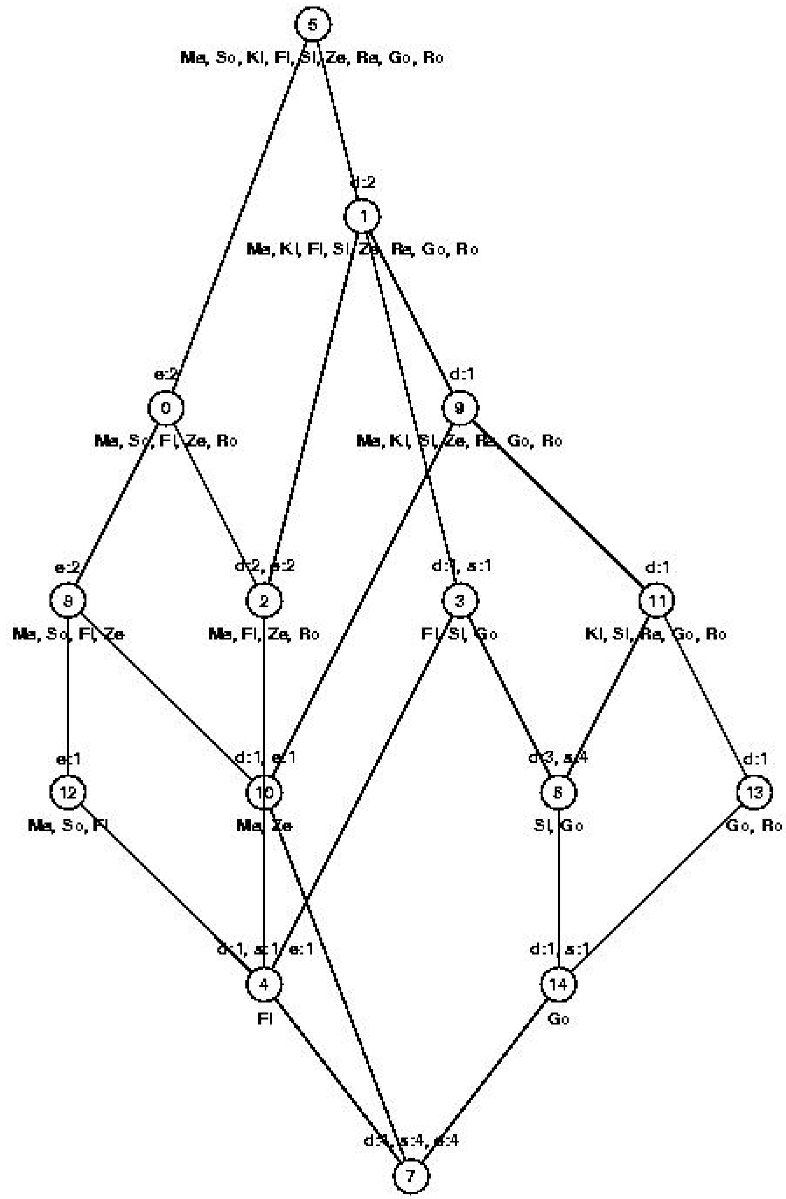
α-řez pro α = 0.7

	d	s	e
Ma			x
So			x
Ki	x		
Fl			x
Sl	x	x	
Ze			x
Ra	x		
Go	x	x	
Ro	x		



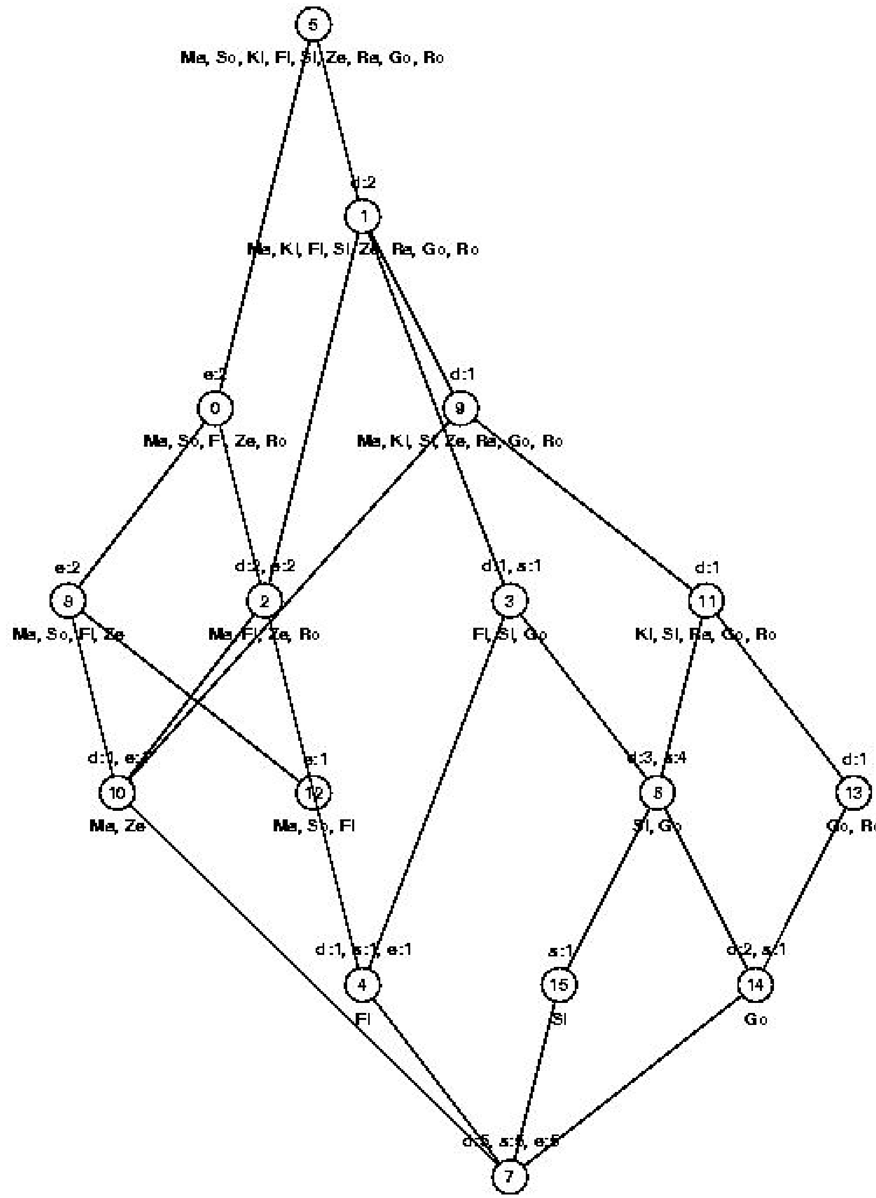
α-řez pro α = 0.8

	d	s	e
Ma			x
So			x
Ki			
Fl			x
Sl		x	
Ze			
Ra			
Go	x	x	
Ro	x		



α-řez pro α = 0.9

	d	s	e
Ma			
So			
Ki			
Fl			
Sl		x	
Ze			
Ra			
Go	x		
Ro			



Závěr

- hierarchická struktura odpovědi
- navigace v odpovědi
- snížení míry subjektivity při volbě škály
- rozeznání významných konceptů (nezávisí na počtu použitých řezů)