

Tato prezentace je součástí wiki-prezentace [Metoda GUHA a systém LISp-Miner](#)

Je dostupná z [této adresy](#)

Verse 1, 6. 7. 2019

Zadávání množin relevantních booleovských atributů v systému LISp-Miner

Jan Rauch

Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze


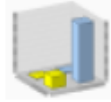

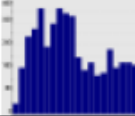




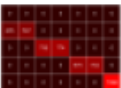

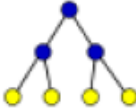

Úvodem

Cíl: Prezentovat hlavní rysy zadávání množin booleovských atributů v systému LISp-Miner.

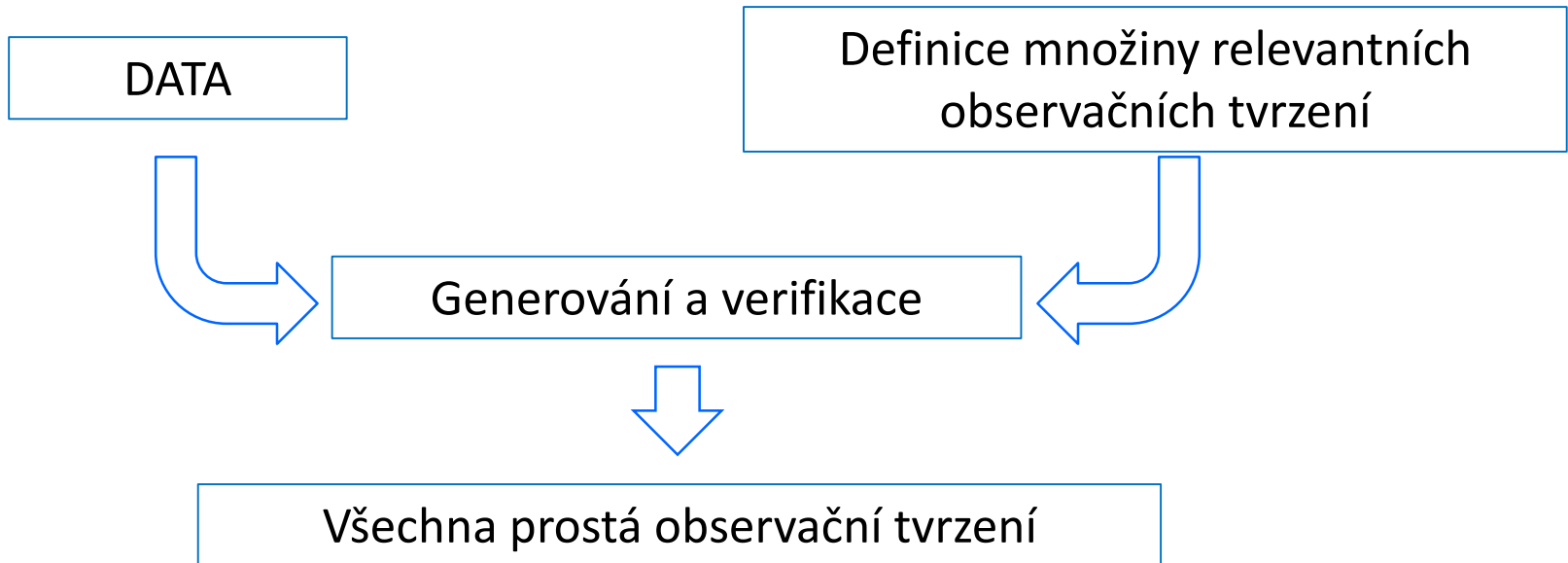
Podrobný popis:

- Kapitola 18 v Rauch, J., Šimůnek, M.: *Dobývání znalostí z databází, LISp-Miner a GUHA*. Oeconomica, VŠE. 2014
- <http://lispminer.vse.cz/wiki/doku.php?id=lmtask:settings:ftcedenthierarchy>
(autor M. Šimůnek)

GUHA procedury systému LIsp-Miner

Základní procedura	Vzor (pattern)	„SD“ procedura
4ft-Miner 	asociační pravidlo	SD4ft-Miner  × 
CF-Miner 	podmíněné rozdělení četností	SDCF-Miner  × 
KL-Miner 	podmíněné K×L frekvence	SDKL-Miner  × 
Ac4ft-Miner 	akční pravidlo	–
ETree-Miner 	rozhodovací nebo explorační strom	–
MCluster-Miner 	shluk (cluster)	–

GUHA procedura



Podstatná část definice množiny relevantních tvrzení:

Definice množiny relevantních booleovských atributů.

Příklad zadání – procedura CF-Miner

CF-Miner



podmíněné
rozdělení četností

Zadání množiny
relevantních cedentů

ATTRIBUTES FOR HISTOGRAM	QUANTIFIERS	CONDITION																																																		
Attributes » DHodnoceni(3 categories)	<table border="1"><thead><tr><th>Type</th><th>Source</th><th>Range</th><th>Rel.</th><th>Value</th><th>Units</th></tr></thead><tbody><tr><td>SUM</td><td>Abs</td><td>all</td><td>>=</td><td>100</td><td>Abs</td></tr><tr><td>S-UP</td><td>%Cond</td><td>all</td><td>>=</td><td>2</td><td>Abs</td></tr></tbody></table>	Type	Source	Range	Rel.	Value	Units	SUM	Abs	all	>=	100	Abs	S-UP	%Cond	all	>=	2	Abs	<table border="1"><thead><tr><th>CONDITION</th><th></th></tr></thead><tbody><tr><td>Pobyt</td><td>Con, 1 - 4</td></tr><tr><td>» PNoci_enum_m (seq), 1 - 2</td><td>B, pos</td></tr><tr><td>» PDenTydne (subset), 1 - 1</td><td>B, pos</td></tr><tr><td>» POsob (seq), 1 - 3</td><td>B, pos</td></tr><tr><td>» POsobonoci_ef5 (seq), 1 - 2</td><td>B, pos</td></tr><tr><td>Host</td><td>Con, 0 - 2</td></tr><tr><td>» HPohlavi (subset), 1 - 1</td><td>B, pos</td></tr><tr><td>» HVek_exp (seq), 1 - 2</td><td>B, pos</td></tr><tr><td>Bydliště</td><td>Con, 0 - 1</td></tr><tr><td>» HCizinec_b (subset), 1 - 1</td><td>B, pos</td></tr><tr><td>» HMesto (subset), 1 - 1</td><td>B, pos</td></tr><tr><td>» HStat (subset), 1 - 1</td><td>B, pos</td></tr><tr><td>Meteo</td><td>Con, 0 - 2</td></tr><tr><td>» MObloha (subset), 1 - 1</td><td>B, pos</td></tr><tr><td>» MTeploata_exp (seq), 1 - 2</td><td>B, pos</td></tr></tbody></table>	CONDITION		Pobyt	Con, 1 - 4	» PNoci_enum_m (seq), 1 - 2	B, pos	» PDenTydne (subset), 1 - 1	B, pos	» POsob (seq), 1 - 3	B, pos	» POsobonoci_ef5 (seq), 1 - 2	B, pos	Host	Con, 0 - 2	» HPohlavi (subset), 1 - 1	B, pos	» HVek_exp (seq), 1 - 2	B, pos	Bydliště	Con, 0 - 1	» HCizinec_b (subset), 1 - 1	B, pos	» HMesto (subset), 1 - 1	B, pos	» HStat (subset), 1 - 1	B, pos	Meteo	Con, 0 - 2	» MObloha (subset), 1 - 1	B, pos	» MTeploata_exp (seq), 1 - 2	B, pos
Type	Source	Range	Rel.	Value	Units																																															
SUM	Abs	all	>=	100	Abs																																															
S-UP	%Cond	all	>=	2	Abs																																															
CONDITION																																																				
Pobyt	Con, 1 - 4																																																			
» PNoci_enum_m (seq), 1 - 2	B, pos																																																			
» PDenTydne (subset), 1 - 1	B, pos																																																			
» POsob (seq), 1 - 3	B, pos																																																			
» POsobonoci_ef5 (seq), 1 - 2	B, pos																																																			
Host	Con, 0 - 2																																																			
» HPohlavi (subset), 1 - 1	B, pos																																																			
» HVek_exp (seq), 1 - 2	B, pos																																																			
Bydliště	Con, 0 - 1																																																			
» HCizinec_b (subset), 1 - 1	B, pos																																																			
» HMesto (subset), 1 - 1	B, pos																																																			
» HStat (subset), 1 - 1	B, pos																																																			
Meteo	Con, 0 - 2																																																			
» MObloha (subset), 1 - 1	B, pos																																																			
» MTeploata_exp (seq), 1 - 2	B, pos																																																			
Aggregate function	Generation information Status: Solved, 15 run(s) Mode: Standard	Total length: 1 - 9																																																		
Type: Count(*) Attribute: -																																																				

Booleovské atributy a jejich množiny v systému LISP-Miner

- Booleovský atribut v GUHA proceduře:
 - Podmínka
 - Antecedent
 - Sukcedent

} definuje množinu objektů (řádků matice)
- Obecný název pro booleovský atribut v GUHA proceduře: **cedent**
- Zajímá nás:
 - Syntaxe cedentu
 - Zadání množiny relevantních cedentů

Cedent a dílčí cedenty (1)

Cedent

$$\varphi = \varphi_1 \wedge \varphi_2 \wedge \dots \wedge \varphi_k$$

Dílčí cedent

Dílčí cedent

Dílčí cedent

$$\varphi_i = \lambda_{i,1} \wedge \dots \wedge \lambda_{i,k_i}$$

konjunkce literálů

literál

literál

Cedent a dílčí cedenty (2)

Cedent

$$\varphi = \varphi_1 \wedge \varphi_2 \wedge \dots \wedge \varphi_k$$

Dílčí cedent

Dílčí cedent

Dílčí cedent

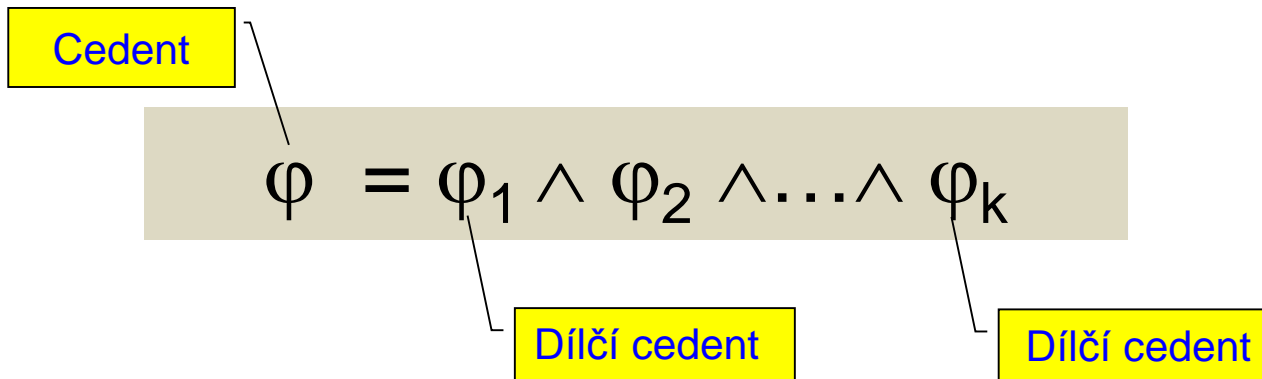
$$\varphi_i = \lambda_{i,1} \vee \dots \vee \lambda_{i,k_i}$$

disjunkce literálů

literál

literál

Zadání množiny relevantních cedentů



Množina relevantních cedentů je dána :

- Zadáním množiny relevantních dílčích cedentů
- ...
- Zadáním množiny relevantních dílčích cedentů
- Minimální a maximální délkou

Zadání množiny relevantních cedentů - příklad

CONDITION

Pobyt	Con, 1 - 4
» PNoci_exp (seq), 1 - 2	B, pos
» PDenTydne (subset), 1 - 1	B, pos
» POsob (seq), 1 - 2	B, pos
» POsobonoci_ef5 (seq), 1 - 2	B, pos
Host	Con, 0 - 2
» HPohlavi (subset), 1 - 1	B, pos
» HVek_exp (seq), 1 - 3	R, pos
Bydliště	Con, 0 - 3
» HCizinec_b (subset), 1 - 1	R, pos
» HMesto (subset), 1 - 1	R, pos
» HStat (subset), 1 - 1	R, pos
Meteo	Con, 0 - 2
» MObloha (subset), 1 - 1	B, pos
» MTeplota_exp (seq), 1 - 2	B, pos

Total length: 1 - 6

Zadání množiny relevantních cedentů

Zadání množin relevantních
dílků cedentů

Zadání minimální a maximální délky cedentu

$$6 \leq 4+2+3+2$$

Zadání množiny relevantních cedentů - příklad

4ft Condition Settings

Basic parameters

Total length: 1 - 9 Edit

Comment: -

List of partial cedents settings

	Number of attributes	Length
Pobyt	4	1 - 4
Host	2	0 - 2
Bydliště	3	0 - 1
Meteo	2	0 - 2

V cedentu se musí vyskytovat alespoň jeden atribut ze skupiny Pobyt

V cedentu se může, ale nemusí vyskytovat atribut ze skupin Host, Bydliště, Meteo

Detail Add Del Up Down Copy Paste Import

Close

Zadání množiny relevantních dílčích cedentů - příklad

4ft Condition Partial cedent Settings

Basic parameters

Name: Host/Bydliště

Min. length: 0 Max. length: 3 Literals boolean operation type: Conjunction

Comment: -

Literals Settings

Underlying attribute	Categories	X-cat	Coefficient type	Length	+/-	B/R	Class of equiv.
HPohlavi	2	No	Subsets	1 - 1	pos	Basic	-
HVek_exp	4	No	Subsets	1 - 1	pos	Remaining	-
HCizinec_b	2	No	Subsets	1 - 1	pos	Remaining	Bydliště
HMesto	28	No	Subsets	1 - 1	pos	Remaining	Bydliště
HStat	5	No	Subsets	1 - 1	pos	Remaining	Bydliště

Buttons: Literal, Coefficient, Eq. Class, Add, Del, Up, Down

Close

12

Minimální a maximální délka

konjunkce / disjunkce

Relevantní dílčí cedent obsahuje nejméně jeden Basic atribut

Relevantní dílčí cedent obsahuje maximálně jeden atribut z každé třídy ekvivalence

Zadání množiny relevantních literálů

Dána:

- Jedním ze sedmi typů literálu
- Minimální a maximální délkou literálu
- Jednou z možností týkající se pozitivních a negativních literálů
 - generovat pouze pozitivní literály
 - generovat pouze negativní literály
 - generovat pozitivní i negativní literály

Zadání množiny relevantních literálů - příklad

Pouze $A(\alpha)$

Pouze $\neg A(\alpha)$

Jak $A(\alpha)$ tak i $\neg A(\alpha)$

Typ α

Nejméně jeden Basic atribut

Minimální a maximální délka α

Literal

Attribute: A

Literal type

- Basic
- Remaining

Gate type

- Positive
- Negative
- Both

Coefficient type

Subset

Coefficient length

Min. length: 1

Max. length: 1

Category

Comment

-

OK Cancel

Typy koeficientů

Coefficient type

Subsets

Subsets

One category

Sequences

Cyclical sequences

Cut

Left cuts

Right cuts

Koeficient *Subsets* – příklad

Atribut A s možnými hodnotami (t.j. kategoriemi) 1, 2, 3, 4, 5

Literály s koeficienty Subset (1 – 3)

A(1), A(2), A(3), A(4), A(5)

A(1, 2), A(1, 3), A(1, 4), A(1, 5)

A(2, 3), A(2, 4), A(2, 5)

A(3, 4), A(3, 5)

A(4, 5)

A(1, 2, 3), A(1, 2, 4), A(1, 2, 5)

A(2, 3, 4), A(2, 3, 5)

A(3, 4, 5)

Literal

Attribute: A

Coefficient type
Subset

Coefficient length
Min. length: 1 Max. length: 3

Category

Koeficient *One category* – příklad

4ft Literal settings

Attribute: HCizinec_b

Literal type

Basic
 Remaining

Gate type

Positive
 Negative
 Both

Coefficient type
One category

Coefficient length

Min. length:	Max. length:
1	1

Hide coefficient value in literals

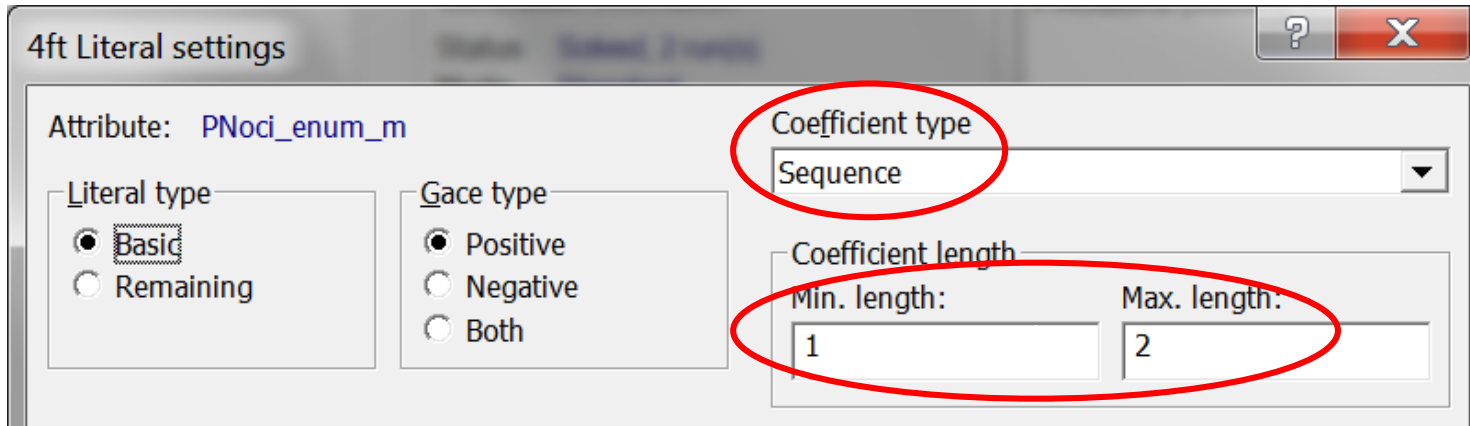
Category
ano

Comment
-

OK Cancel Categories frequency

Zajímají nás jenom cizinci

Koeficient *Sequences* – příklad



10 kategorií : 1, 2, ⟨3;6⟩, 7, ⟨8;13⟩, 14, ⟨15;20⟩, 21, ⟨22;27⟩, 28

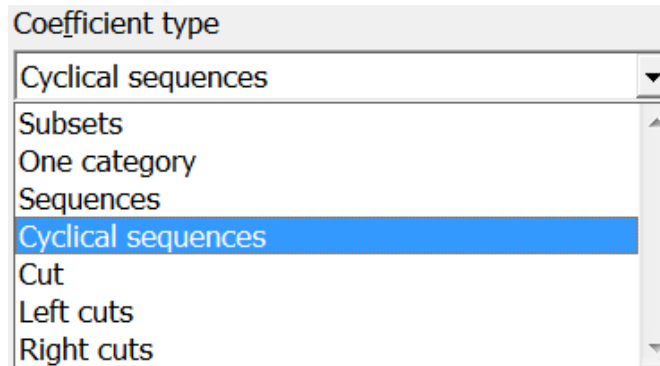
PNoci(1)	● ● ● ● ● ● ● ● ● ●
PNoci(2)	● ● ● ● ● ● ● ● ● ●
Pnoci(⟨3;6⟩)	● ● ● ● ● ● ● ● ● ●
PNoci(7)	● ● ● ● ● ● ● ● ● ●
PNoci(⟨8;13⟩)	● ● ● ● ● ● ● ● ● ●
PNoci(14)	● ● ● ● ● ● ● ● ● ●
PNoci(⟨15;20⟩)	● ● ● ● ● ● ● ● ● ●
PNoci(21)	● ● ● ● ● ● ● ● ● ●
PNoci(⟨22;27⟩)	● ● ● ● ● ● ● ● ● ●
PNoci(28)	● ● ● ● ● ● ● ● ● ●

PNoci(1,2)	● ● ● ● ● ● ● ● ● ●
PNoci(2,⟨3;6⟩)	● ● ● ● ● ● ● ● ● ●
PNoci(⟨3;6⟩,7)	● ● ● ● ● ● ● ● ● ●
PNoci(7,⟨8;13⟩)	● ● ● ● ● ● ● ● ● ●
PNoci(⟨8;13⟩,14)	● ● ● ● ● ● ● ● ● ●
PNoci(14,⟨15;20⟩)	● ● ● ● ● ● ● ● ● ●
PNoci(⟨15;20⟩,21)	● ● ● ● ● ● ● ● ● ●
PNoci(21, ⟨22;27⟩)	● ● ● ● ● ● ● ● ● ●
PNoci(⟨22;27⟩,28)	● ● ● ● ● ● ● ● ● ●

Koeficient *Cyclical sequences* – příklad

Atribut: *Den týdne* s hodnotami po, út, st, čt, pá, so, ne

zadání koeficientu:



generované pozitivní literály délky 3:

Den týdne (po, út, st)

Den týdne (út, st, čt)

Den týdne (st, čt, pá)

Den týdne (čt, pá, so)

Den týdne (pá, so, ne)

Den týdne (so, ne, po)

Den týdne (ne, po, út)

Koeficient *Left cuts* – příklad

4ft Literal settings

Attribute: Hvek

Literal type
 Basic
 Remaining

Case type
 Positive
 Negative
 Both

Coefficient type
Left cuts

Coefficient length
Min. length: 1
Max. length: 4

Hide coefficient value in literals

Category

Comment
-

OK Cancel Categories frequency

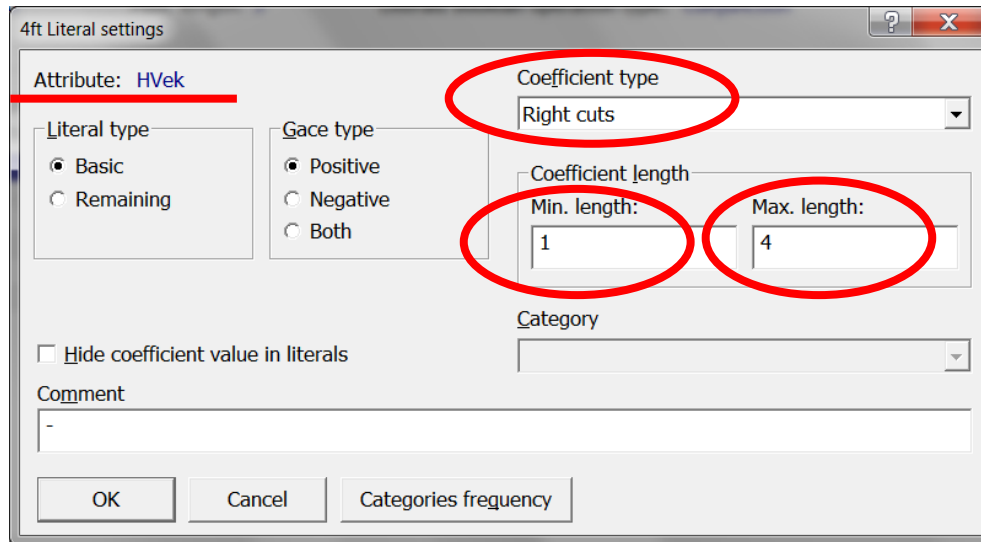
18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,....., 75, 76, 77, 78, 79, 80, 81, 82

18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,....., 75, 76, 77, 78, 79, 80, 81, 82

18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,....., 75, 76, 77, 78, 79, 80, 81, 82

18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,....., 75, 76, 77, 78, 79, 80, 81, 82

Koeficient *Right cuts* – příklad



18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,....., 75, 76, 77, 78, 79, 80, 81, 82

18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,....., 75, 76, 77, 78, 79, 80, 81, 82

18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,....., 75, 76, 77, 78, 79, 80, 81, 82

18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,....., 75, 76, 77, 78, 79, 80, 81, 82

Koeficient *Cuts* – příklad

Cut – generují se
Left cut i *Right Cut*

4ft Literal settings

Attribute: **HVeK**

Literal type: Basic Remaining

Gace type: Positive Negative Both

Coefficient type: **Cut**

Coefficient length: Min. length: 1 Max. length: 4

Category: []

Hide coefficient value in literals

Comment: -

OK Cancel Categories frequency

4ft Literal settings

Attribute: **HVeK**

Literal type: Basic Remaining

Gace type: Positive Negative Both

Coefficient type: **Left cuts**

Coefficient length: Min. length: 1 Max. length: 4

Category: []

Hide coefficient value in literals

Comment: -

OK Cancel Categories frequency

4ft Literal settings

Attribute: **HVeK**

Literal type: Basic Remaining

Gace type: Positive Negative Both

Coefficient type: **Right cuts**

Coefficient length: Min. length: 1 Max. length: 4

Category: []

Hide coefficient value in literals

Comment: -

OK Cancel Categories frequency