

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

Je dostupná z [této adresy](#)

Verse 30. 7. 2019

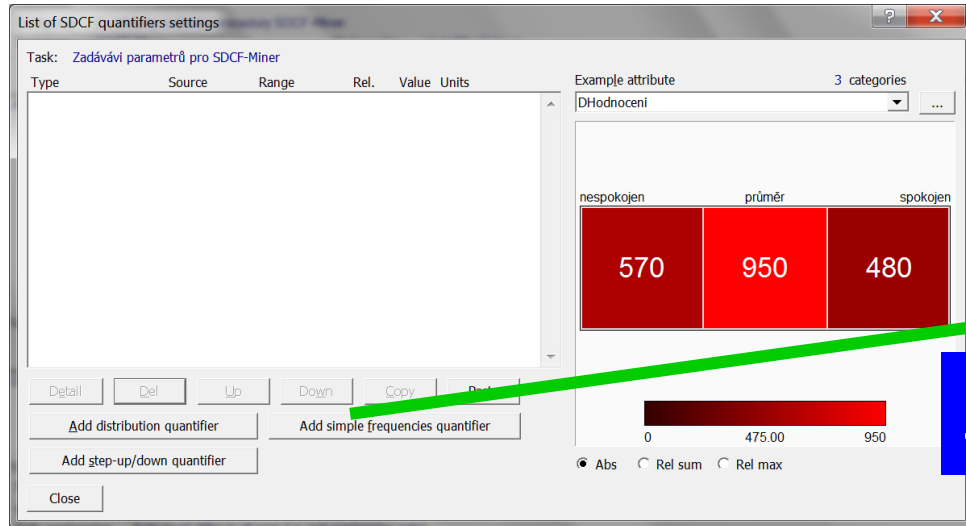
# Zadávání jednoduchých frekvenčních SDCF-kvantifikátorů pro proceduru SDCF-Miner

Jan Rauch

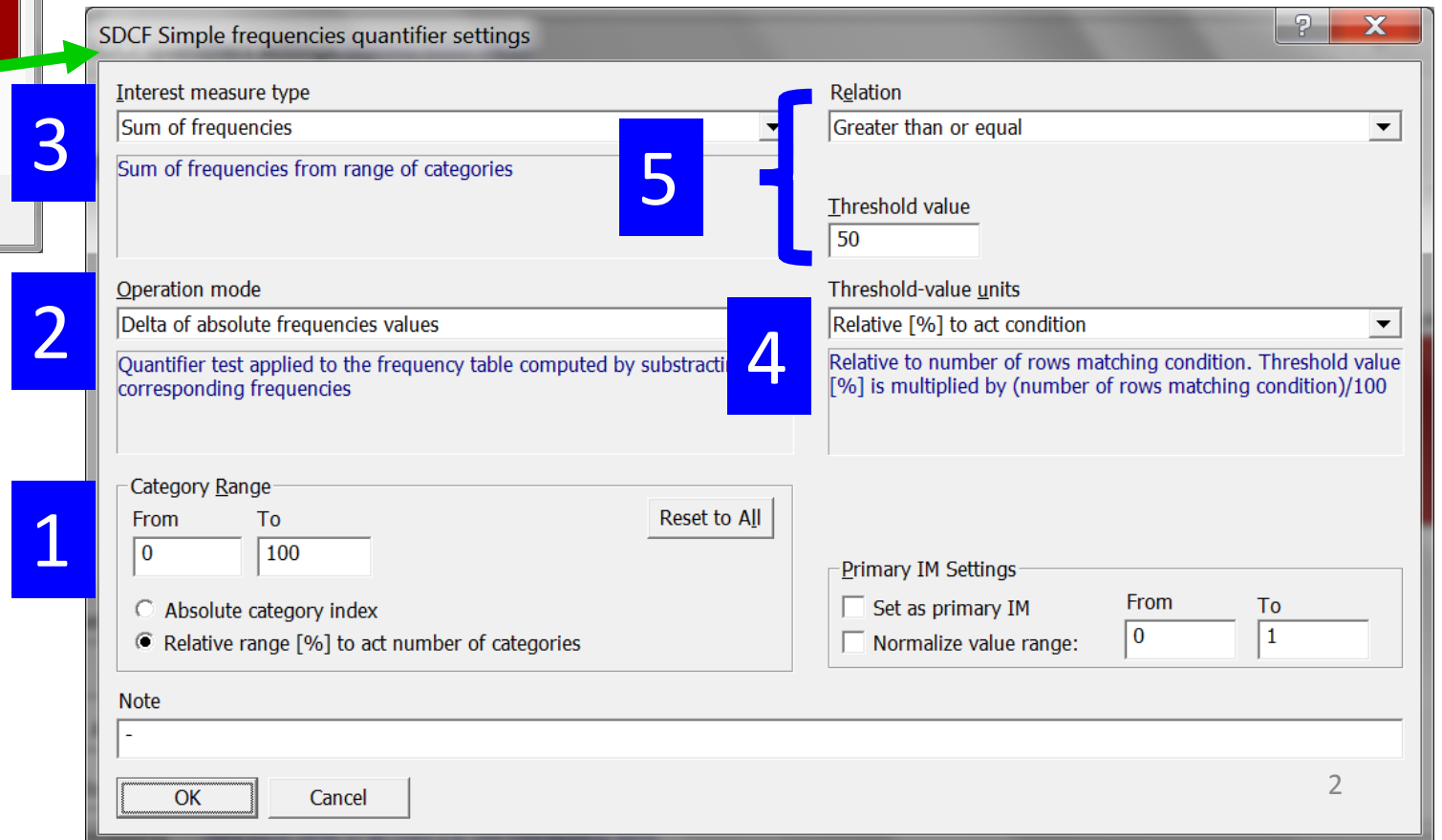
Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

# Start tlačítkem Add simple frequencies quantifier



- Definice se provádí zadáním parametrů 1 - 5.
- Pro všechny parametry se nabízejí defaultní hodnoty.
- Podmínka kvantifikátoru je definována parametrem 5.



# Výchozí SDCF-tabulka

Jednoduchý frekvenční SDCF-kvantifikátor se aplikuje na SDCF-tabulku  $CF(A, \alpha, \beta, \chi, M)$

kategorie atributu A	$a_1$	...	$a_p$	...	$a_q$	...	$a_K$	$\Sigma$
absolutní četnosti v matici $M/\chi \wedge \alpha$	$n_{\alpha,1}$	...	$n_{\alpha,p}$	...	$n_{\alpha,q}$	...	$n_{\alpha,K}$	$n_\alpha$
absolutní četnosti v matici $M/\chi \wedge \beta$	$n_{\beta,1}$	...	$n_{\beta,p}$	...	$n_{\beta,q}$	...	$n_{\beta,K}$	$n_\beta$
absolutní četnosti v matici $M/\chi$	$n_1$	...	$n_p$	...	$n_q$	...	$n_K$	$n$
absolutní četnosti v matici M	$m_1$	...	$m_p$	...	$m_q$	...	$m_K$	$n_T$

# 1 - Category Range

- Category Range určuje výsek SDCF-tabulky, na který budou aplikovány ostatní parametry.
- Provádí se způsobem popsáním [zde](#).
- Výsek SDCF-tabulky týkající se  $K$  kategorií je dán dvojicí  $\langle p, q \rangle$  celých čísel splňujících  $1 \leq p \leq q \leq K$ .
- Výsek  $\langle p, q \rangle$  SDCF-tabulky zahrnuje frekvence pro kategorie  $a_p, \dots, a_q$ .

Category Range

From  To

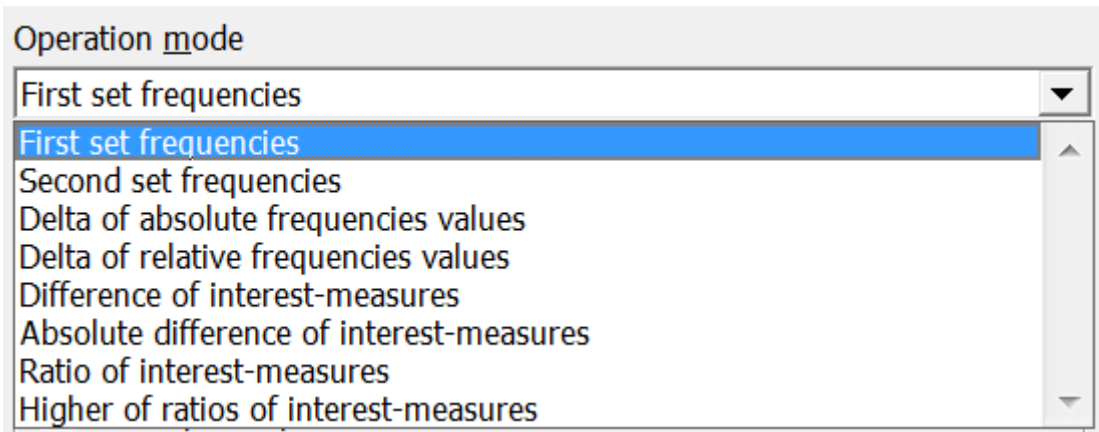
Absolute category index

Relative range [%] to act number of categories

kategorie atributu A	$a_1$	...	$a_p$	...	$a_q$	...	$a_K$	$\Sigma$
absolutní četnosti v matici $M/\chi \wedge \alpha$	$n_{\alpha,1}$	...	$n_{\alpha,p}$	...	$n_{\alpha,q}$	...	$n_{\alpha,K}$	$n_\alpha$
absolutní četnosti v matici $M/\chi \wedge \beta$	$n_{\beta,1}$	...	$n_{\beta,p}$	...	$n_{\beta,q}$	...	$n_{\beta,K}$	$n_\beta$
absolutní četnosti v matici $M/\chi$	$n_1$	...	$n_p$	...	$n_q$	...	$n_K$	$n$
absolutní četnosti v matici M	$m_1$	...	$m_p$	...	$m_q$	...	$m_K$	$n_T$

## 2 - Operation mode

Vybírá se jeden z operačních módů nabízených v menu *Operation mode*. Ten určuje, jakým způsobem bude aplikována vybraná míra zajímavosti na SDCF-tabulku.



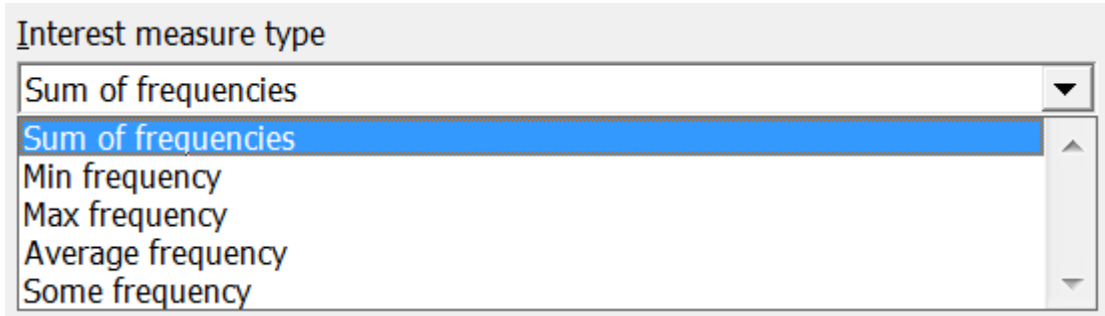
Pro zbývající čtyři nabízené módy se vybraná míra zajímavosti aplikuje zvlášť na části SDCF tabulky pro atributy  $\alpha$  a  $\beta$ . Poté se výsledky dále zpracují.

Pro první čtyři nabízené Operation mode se nejprve vytvoří vektor  $\langle z_p, \dots, z_q \rangle$  takto:

- pro *First set frequencies*  $\langle z_p, \dots, z_q \rangle = \langle n_{\alpha,p}, \dots, n_{\alpha,q} \rangle$
- pro *Second set frequencies*  $\langle z_p, \dots, z_q \rangle = \langle n_{\beta,p}, \dots, n_{\beta,q} \rangle$
- pro *Delta of absolute frequencies values*  
 $\langle z_p, \dots, z_q \rangle = \langle |n_{\alpha,p} - n_{\beta,p}|, \dots, |n_{\alpha,q} - n_{\beta,q}| \rangle$
- pro *Delta of relative frequencies values*  
 $\langle z_p, \dots, z_q \rangle = \langle \left| \frac{n_{\alpha,p}}{n_\alpha} - \frac{n_{\beta,p}}{n_\beta} \right|, \dots, \left| \frac{n_{\alpha,q}}{n_\alpha} - \frac{n_{\beta,q}}{n_\beta} \right| \rangle.$

## 3 - Interest measure type (A)

Vybírá se jedna z měr zajímavosti nabízených v menu *Interest measure type*.



Pro výpočet hodnoty  $IM$  pro první čtyři operační módy

- *First set frequencies*
- *Second set frequencies*
- *Delta of absolute frequencies values,*
- *Delta of relative frequencies values*

se použije vektor  $\langle z_p, \dots, z_q \rangle$  dle parametru **Operation mode**.

Na základě Interest measure type se určí hodnota  $IM$  pro porovnání s prahem dle relace zadané parametrem 5. Platí:

- pro *Sum of frequencies* se spočítá hodnota  $IM = \sum_{i=p}^q z_i$
- pro *Min frequency* se spočítá hodnota  $IM = \min_i \{z_i | i = p, \dots, q\}$
- pro *Max frequency* se spočítá hodnota  $IM = \max_i \{z_i | i = p, \dots, q\}$
- pro *Average frequency* se spočítá  $IM = \frac{\sum_{i=p}^q z_i}{(q-p+1)}$
- pro *Some frequency* se zjišťuje, zda alespoň jedno  $z_i$  pro  $i = p, \dots, q$  splňuje zadanou nerovnost.

### 3 - Interest measure type (B)

Při výpočtu hodnoty  $IM$  pro zbývající čtyři operační módy se nejprve se vypočtou hodnoty  $IM_\alpha$  a  $IM_\beta$  pro vybraný *Interest measure type*:

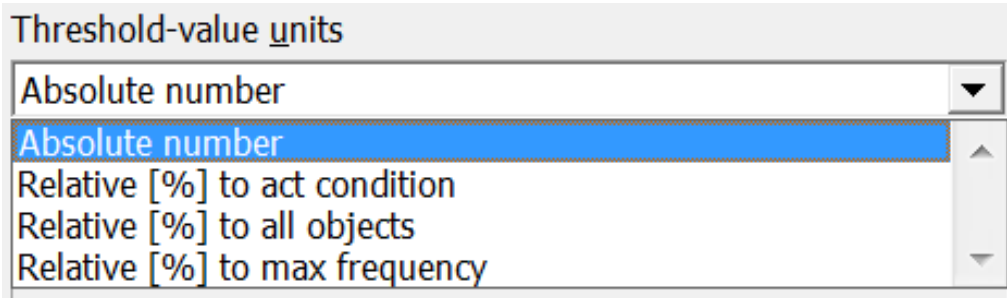
Interest measure type	Hodnota $IM_\alpha$	Hodnota $IM_\beta$
Sum of frequencies	$\sum_{i=p}^q n_{\alpha,i}$	$\sum_{i=p}^q n_{\beta,i}$
Min frequency	$\min_i \{n_{\alpha,i}   i = p, \dots, q\}$	$\min_i \{n_{\beta,i}   i = p, \dots, q\}$
Max frequency	$\max_i \{n_{\alpha,i}   i = p, \dots, q\}$	$\max_i \{n_{\beta,i}   i = p, \dots, q\}$
Average frequency	$\frac{\sum_{i=p}^q n_{\alpha,i}}{(q - p + 1)}$	$\frac{\sum_{i=p}^q n_{\beta,i}}{(q - p + 1)}$
Some frequency	Výpočet není k dispozici, není rozumná definice	

Poté se vypočte hodnota  $IM$  v závislosti na vybraném *Operation mode*:

Operation mode	$IM$
Difference of interest-measures	$IM_\alpha - IM_\beta$
Absolute difference of interest-measures	$ IM_\alpha - IM_\beta $
Ratio of interest-measures	$\frac{IM_\alpha}{IM_\beta}$
Higher of ratios of interest-measures	$\max\left\{\frac{IM_\alpha}{IM_\beta}, \frac{IM_\beta}{IM_\alpha}\right\}$

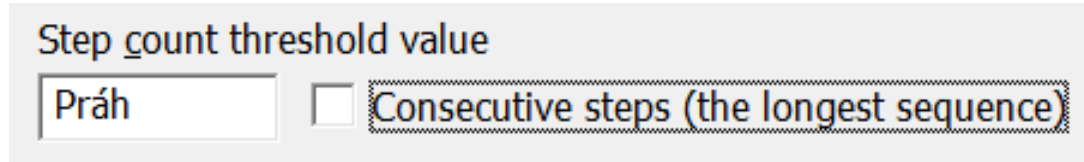
## 4 - Threshold-value units

Vybírá se jedna z voleb nabízených v menu *Threshold-value units*



Threshold-value units

- Absolute number
- Absolute number
- Relative [%] to act condition
- Relative [%] to all objects
- Relative [%] to max frequency



Step count threshold value

Práh  Consecutive steps (the longest sequence)

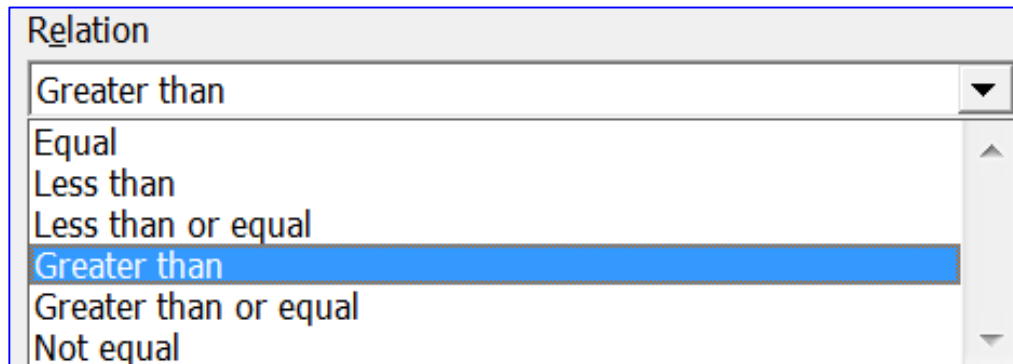
Na základě této volby se určí jednotky, ve kterých je hodnota *Práh* zadaná v poli *Threshold value*. To znamená, že se před porovnáním dle parametru 5 zadaná hodnota *Práh* vynásobí vhodným číslem dle volby v menu. Výslednou hodnotu pro porovnání značíme *PráhAkt*. Platí:

- pro *Absolute number*  $PráhAkt = Práh$
- pro *Relative [%] to act condition*  $PráhAkt = \frac{100}{n} * Práh$
- pro *Relative [%] to all objects*  $PráhAkt = \frac{100}{n_T} * Práh$
- pro *Relative [%] to max frequency*  $PráhAkt = \frac{100}{Max} * Práh$  kde  $Max = \max_i \{m_i | i = 1, \dots, K\}$



## 5 - Relation x Threshold value

Na základě volby v nabídce *Relation* se vybere relace, která se použije pro porovnání hodnoty *IM* vypočtené dle parametru 3 - *Inte*nterest measure type s hodnotou *PráhAkt* vypočítanou dle parametru 4 - *Threshold-value units* .



Relation

Greater than

Equal

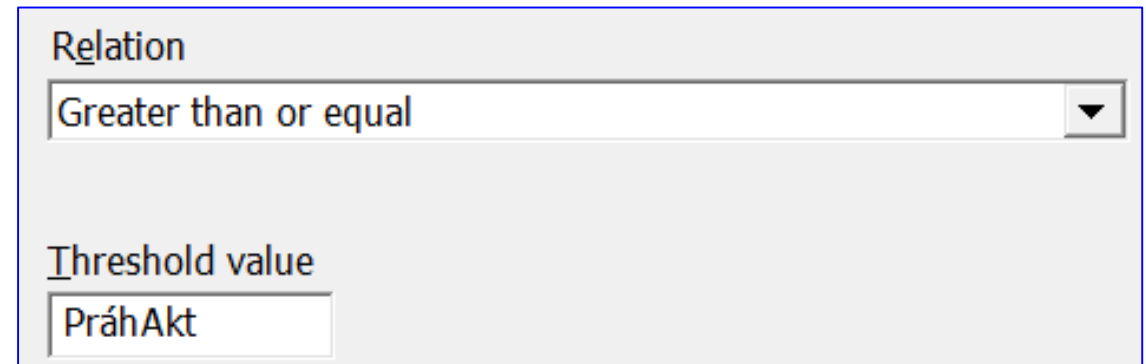
Less than

Less than or equal

Greater than

Greater than or equal

Not equal



Relation

Greater than or equal

Threshold value

PráhAkt

Platnost vybrané relace je považována za podmínku definující SDCF-kvantifikátor zadaný parametry 1 až 5.