

Two Lists of Terms Similarity Percentage Formula

The formula:

outputs the similarity percentage between two lists of terms.

$A \ni t$

$(B \ni t \rightarrow b = 1) \wedge (B \not\ni t \rightarrow b = 0)$

$$Sim = \frac{\sum_{i=0}^n b(n - |i_{t_i} - j_{t_i}|)}{n^2}$$

Where:

A is the first list of term

B is the second list of the term

t is the term

b is the Boolean coefficient (t in B)

n is the number of terms being compared

i_t is the index of the term in the first list

j_t is the index of the term in the second list

The lists index weight sum formula:

outputs the index weight or location weight of a term relative to both lists.

$A \ni t$

$(B \ni t \rightarrow b = 1) \wedge (B \not\ni t \rightarrow b = 0)$

$$S = \sum_{i=0}^n b(n - |i_{t_i} - j_{t_i}|)$$

Where:

S is the sum of the index (location) weight of the terms in list A

A is the first list of term

B is the second list of the term

t is the term

b is the Boolean coefficient (t is in B)

n is the number of terms being compared

i_t is the index of the term in the first list

j_t is the index of the term in the second list

The lists index weight formula:

$A \ni t$

$(B \ni t \rightarrow b = 1) \wedge (B \not\ni t \rightarrow b = 0)$

$$w_{t_i} = b(n - |i_{t_i} - j_{t_i}|)$$

Where:

w_{t_i} is the index weight of the terms t_i in list A

A is the first list of term

B is the second list of the term

t is the term

b is the Boolean coefficient (t is in B)

n is the number of terms being compared

i_t is the index of the term in the first list

j_t is the index of the term in the second list