

# Modelo do espaço vetorial

Representação interna da consulta (Q) e dos elementos de informação (D<sub>i</sub>) em vetores de  $n$  dimensões.

$$Q = \{q_1, q_2, \dots, q_n\}$$

$$D_i = \{p_{1i}, p_{2i}, \dots, p_{ni}\}$$

onde:  $p_{ij}$  = peso da palavra chave  $K_j$  no elemento de informação  $D_i$

$$p_{ij} = tf_{ij} \times idf_j$$

onde:

$tf_{ij}$  = frequência de ocorrência da palavra chave  $K_j$  no elemento de informação  $D_i$

$idf_j$  = frequência de ocorrência da palavra chave  $K_j$  em toda a fonte de informação

$$idf_j = \log (N/n_j)$$

onde:

$N$  = número de documentos na fonte de informação

$n_j$  = número de documentos na fonte de informação com ocorrência da palavra chave  $K_j$

$$\text{similarity}(Q, D_i) = \sum_{i=1, n} q_i \times p_{ij} / ( \sqrt{\sum_{i=1, n} q_i^2} \times \sqrt{\sum_{i=1, n} p_{ij}^2} )$$