

VSM Is a Framework

- Represent a doc/query by a term vector
 - **Term**: basic concept, e.g., word or phrase
 - Each term defines one dimension
 - N terms define an **N-dimensional space**
 - **Query** vector: $\mathbf{q}=(x_1, \dots x_N)$, $x_i \in \mathbb{R}$ is query term weight
 - **Doc** vector: $\mathbf{d}=(y_1, \dots y_N)$, $y_j \in \mathbb{R}$ is doc term weight
- $\text{relevance}(\mathbf{q}, \mathbf{d}) \propto \text{similarity}(\mathbf{q}, \mathbf{d}) = f(\mathbf{q}, \mathbf{d})$