

# Generative model of texts

## Probabilistic Latent Semantic Analysis (PLSA):

$$p(w|d) = \sum_{t \in T} p(w|t, d) p(t|d) = \sum_{t \in T} p(w|t) p(t|d)$$



*Law of total probability*

$$p(w) = \sum_{t \in T} p(w|t) p(t)$$

## Notation:

- $w$  – word
- $d$  – document
- $t$  – topic