**Topic Modelling** is the task of using unsupervised learning to extract the main topics (represented as a set of words) that occur in a collection of documents.

LDA: - Topics per document and words per topic.

* Each document is modeled as a multinomial distribution of topics and each topic is modeled as a multinomial distribution of words.

Document🡪 Multinomial distribution🡪 topic

Topic 🡪 Multinomial distribution🡪 words

Multinomial distribution🡺 Binomial distribution

* LDA assumes that every chunk of text we feed into it will contain words that are somehow related. Therefore, choosing the right corpus of data is crucial.
* It also assumes documents are produced from a mixture of topics. Those topics then generate words based on their probability distribution.