

TF-IDF Vectorization

TF-IDF stands for Term Frequency - Inverse Document Frequency.
It reflects how important a word is to a document in a collection.

TF-IDF is widely used in text mining and information retrieval.
It gives higher scores to terms that appear often in a document
but rarely in the rest of the corpus.

BM25 can be seen as an improvement of TF-IDF for ranking purposes.