



|  |  |
| --- | --- |
| fit\_on\_sequences(sequences) | Updates internal vocabulary based on a list of sequences |
| fit\_on\_texts(texts) | Updates internal vocabulary based on a list of texts. |
| get\_config() | Returns the tokenizer configuration as Python dictionary. The word count dictionaries used by the tokenizer get serialized into plain JSON, so that the configuration can be read by other projects |
| sequences\_to\_matrix(sequences, mode='binary') | Converts a list of sequences into a Numpy matrix. |
| sequences\_to\_texts( sequences) | Transforms each sequence into a list of text. |
| sequences\_to\_texts\_generator(sequences) | Transforms each sequence in sequences to a list of texts(strings). |
| texts\_to\_matrix(texts, mode='binary') | Convert a list of texts to a Numpy matrix. |
| texts\_to\_sequences(texts) | Transforms each text in texts to a sequence of integers. |