"We thus define the tidy text format as being a table",

"with one-token-per-row. A token is a meaningful unit of text, such as a word",

"that we are interested in using for analysis, and tokenization is",

"the process of splitting text into tokens. This one-token-per-row",

"structure is in contrast to the ways text is often stored in",

"current analyses, perhaps as strings or in a document-term matrix.",

"For tidy text mining, the token that is stored in each row is",

"most often a single word, but can also be an n-gram, sentence, or",

"paragraph. In the tidytext package, we provide functionality to",

"tokenize by commonly used units of text like these and convert",

"to a one-term-per-row format."