

Here is a sample document for testing the Text Cleaning and Pre-Processing Tool. This text includes various features: punctuation like commas, periods, and question marks? Yes! It also has numbers (e.g., 1, 2, 345) and mixed case words (like "Cleaning" and "processing").

We are hoping this tool works really well. The NLTK library is powerful. Stop words such as 'the', 'is', 'and', 'are' should be removed. Words like 'running', 'runs', and 'ran' could be stemmed to 'run'. Similarly, 'better' and 'best' might be lemmatized to 'good'.

Check out our website: <https://www.google.com/search?q=https://www.example-tool.com> or email us at contact@example-tool.com. This <html> tag should also be removed.

This is the end of the test file. Good luck!