

Hello World!

This is a simple PDF document to test the PDF reading functionality in Java.

It contains a few different words, including unique ones like pineapple, quantum, and zephyr.

Let's see if your Java code can handle this and extract the words properly.

Hello, hello, HELLO! How are you today? I hope you are doing well. Today is a beautiful day. Let's enjoy the sunshine and the fresh air.

The quick brown fox jumps over the lazy dog. The quick brown fox jumped over the lazy dogs. This is an example of a pangram that contains every letter of the alphabet.

12345 is a number, not a word. However, "word123" and "123word" might be treated as words depending on the implementation.

Special characters such as @, #, \$, and % are not words but symbols. Should they be included?

Multiple spaces should not affect the outcome. Tabs also should be treated the same as spaces.

Can you find this: example, Example, EXAMPLE? They are the same word normalized but appear differently in the text.