NLP Task: Tokenization

Tokenization is the process of breaking a text into smaller pieces. These pieces can be sentences or words.  
  
In this example, we use Python's nltk library to:  
1. Break a paragraph into sentences (Sentence Tokenization)  
2. Break the same paragraph into words (Word Tokenization)

# Python Code

import nltk  
nltk.download('punkt')  
from nltk.tokenize import word\_tokenize, sent\_tokenize  
  
text = "Natural Language Processing is fun. Let's learn it step by step!"  
  
# Sentence Tokenization  
sentences = sent\_tokenize(text)  
print("Sentence Tokenization:")  
print(sentences)  
  
# Word Tokenization  
words = word\_tokenize(text)  
print("\nWord Tokenization:")  
print(words)

# Sample Output

Sentence Tokenization:  
['Natural Language Processing is fun.', "Let's learn it step by step!"]  
  
Word Tokenization:  
['Natural', 'Language', 'Processing', 'is', 'fun', '.', 'Let', "'s", 'learn', 'it', 'step', 'by', 'step', '!']