NLTK\_Cheatsheet.docx

Table of Contents

[Tweet Tokenizer 1](#_Toc50966601)

# Tweet Tokenizer

from nltk.tokenize import TweetTokenizer # module for tokenizing strings

print()

print('\033[92m' + tweet2)

print('\033[94m')

# instantiate tokenizer class

tokenizer = TweetTokenizer(preserve\_case=False, strip\_handles=True,

reduce\_len=True)

# tokenize tweets

tweet\_tokens = tokenizer.tokenize(tweet2)

print()

print('Tokenized string:')

print(tweet\_tokens)

Output:

**My beautiful sunflowers on a sunny Friday morning off :) sunflowers favourites happy Friday off…**

**Tokenized string:**

**['my', 'beautiful', 'sunflowers', 'on', 'a', 'sunny', 'friday', 'morning', 'off', ':)', 'sunflowers', 'favourites', 'happy', 'friday', 'off', '…']**