

```
1 import nltk
2 nltk.download('punkt')
3
4 text = "Code is like humor. When you have to explain it, it's bad."
5
6 split_tokens = text.split()
7 print(f"split tokens:\n{split_tokens}")
8 print(f"length of split tokens: {len(split_tokens)}\n")
9
10 nltk_tokens = nltk.word_tokenize(text)
11 print(f"nltk tokens:\n{nltk_tokens}")
12 print(f"length of nltk tokens: {len(nltk_tokens)}\n")

→ split tokens:
['Code', 'is', 'like', 'humor.', 'When', 'you', 'have', 'to', 'explain', 'it,', 'it's', 'bad.']
length of split tokens: 12

nltk tokens:
['Code', 'is', 'like', 'humor', '.', 'When', 'you', 'have', 'to', 'explain', 'it', ',', 'it', "'", 's', 'bad', '.']
length of nltk tokens: 17

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
```